0021-7557/07/83-05/395
Jornal de Pediatria

An evident view of evidence-based practice in perinatal medicine: absence of evidence is not evidence of absence

Uma visão evidente da prática baseada em evidências na medicina perinatal: ausência de evidência não é evidência de ausência

Augusto Sola[1], Fernando Dominguez Dieppa[2], Marta R. Rogido[3]

Abstract

Objective: To provide valuable elements and some humor in this so-called era of "evidence-based practice" with the aim of helping clinicians make better choices in the care they deliver based on evidence, not simply or exclusively based on a randomized clinical trial (RCT) or meta-analysis (which may not be evidence).

Sources: Books and peer-reviewed articles are quoted and listed in the bibliography. Evidence of life, learning from our own mistakes and many other evident facts that support this review are not quoted.

Summary of the findings: 1) "Absence of evidence is not evidence of absence" and "lack of evidence of effect does not mean evidence of no effect". 2) RCTs with "negative" results and those with "positive" results, but without outcomes that matter, often cannot conclude what they conclude. 3) Non-randomized clinical trials and practical trials may be important. 4) Research to prove is different than research to improve. 5) Clinical choice must assess effects on outcomes that matter to patients and their parents. 6) Quantifying adverse outcomes, number needed to damage and to treat is not that simple.

Conclusions: Significant challenges inherent to health service research must be correlated to possible clinical applications using tools to have a more "evident view of evidence-based practice" in perinatal medicine, recalling that absence of evidence is not evidence of absence.

J Pediatr (Rio J). 2007;83(5):395-414: Evidence-based medicine, number needed to treat, randomized trials, outcome variables, treatment effects, critical reading, statistical significance.

Introduction

The non-equivalence of statistical significance and clinical importance has long been recognized; this error of interpretation remains common. A significant result may sometimes not be clinically important. Much worse is the misinterpretation of "nonsignificant" findings. Other common misinterpretations are the confusions between "evidence of no effect" with "no evidence of effect" and "absence of evidence" with "no evidence of absence." All these factors have an impact on the application of results from clinical research into clinical practice. Hence, this review is important for the practice of neonatology and pediatrics.

Several people have educated and inspired us in many or all of the concepts we will share in this review.

Our aim is to summarize some "evident concepts of evidence based practice" in a simple and user-friendly way, with some humor every so often, to remind us that "one of the first symptoms of an approaching nervous breakdown is the belief that one's work is terribly important" (Bertrand Russell). Our hope is that some of the concepts covered in this review will

1. MD. MANA and Atlantic Neonatal Research Institute, Division of Neonatology, Atlantic Health System, Morristown, NJ, USA.
2. MD. Gonzalez Coro University Hospital, La Habana, Cuba.
3. MD. MANA and Atlantic Neonatal Research Institute, Division of Neonatology, Atlantic Health System, Morristown, NJ, USA.

Suggested citation: Sola A, Dieppa FD, Rogido MR. An evident view of evidence-based practice in perinatal medicine: absence of evidence is not evidence of absence. J Pediatr (Rio J). 2007;83(5):395-414.

Manuscript received Jul 06 2007, accepted for publication Jul 26 2007.

doi 10.2223/JPED.1702

Table 1 - Type I and II errors and the power of a test

Type I Error	Type II Error	Power of the test
We conclude that there is a difference when in fact there is none. The associated probability is called alpha.	Concluding that there is no difference when there truly is one in the real world.	The probability of concluding that there is a difference when in fact there is one (usually set at 80%, derived from 1- beta or type II error).

become of value to clinicians for daily practice and delivery of care and, therefore, be of worth for one infant, one day, some place.

Basic concepts of statistical inference

There are many significant flaws in selecting and reporting denominators and statistical tests in the medical literature. Two books describe this with great humor and easy reading: "Bare Essentials of Biostatistics"[1] and "Biomedical Bestiary".[2]

One modern "intellectual disease" in the literature is the inappropriate utilization of statistical significance. Many authors, peer reviewers and editors "die for" statistical significance and for "p < 0.05." However, some of them neither quite understand its real meaning nor fully comprehend if the appropriate statistical method has been applied correctly. W. Castle has said: *"Most researchers use statistics the way a drunkard uses a lamppost: More for support than for illumination"*.

Clinicians and statistical inference are sometimes at odds. Inferential statistics are used to determine the probability of a conclusion based on data analysis being true and to quantify the degree of inaccuracy of the estimate. And then, "we play a game". It is accepted that if a difference can occur more than 5 times in 100, there is too great of a probability that the difference is due to chance alone. On the other hand, if the probability that the difference is due to chance < 5% of the time, we say that the difference "demonstrates" that the two samples are actually different. Now you know what p < 0.05 means, or almost. If not, read again, and keep reading, please. In summary, a p value < 0.05 means that 95% of the time or more the results found are not due to chance. If the likelihood of chance affecting the results is less than 1 in 20, then one might regard the result as significant, as Sir R. Fisher said.

For type I and type II errors and the power of a test, see Table 1 before reading further. When we use a p level of 0.05, we accept that 5% of the time we can be making a type I error. Since experiments are usually done to demonstrate differences, statisticians are often interested in the probability of detecting a true difference. This 5%, however, is not some absolute criterion of truth. If, for example, an effect exists at exactly p = 0.049, it does not suddenly disappear at p = 0.051. Rosnow eloquently said: *"Surely, God loves the 0.06 nearly as much as the 0.05"*.

A frequent error in the quest for a p < 0.05 is using the Student's *t* test when analyzing repeated measures or variables. If the outcome were blood pressure changes over time and four values are reported, making statistical comparisons *within* or *between* groups at those time points, repeating *t* tests for each of the comparisons increases the chance of showing statistically significant difference, when in reality there is none. Imagine flipping a coin one time after another. Let us assume that the first time you get heads (tails also had a prior probability of 50%, you surely agree). As of that time however, the chances for tails are higher in each successive flip. By 3, 4 or 5 tosses, the chances to get tails have increased exponentially to 84-93% (don't waste time with mathematics; just believe us on this one). Similarly, when repeating t test after *t* test, the probability to get a significant p just by chance increases to about 30% by the 5th-6th time, even when no true difference exists. The *t* test is very valid when the means of two groups are compared, but when dealing with more than two groups or with repeated measurements in two or more groups, the *t* test is incorrect and not valid statistically. Be aware of this, and do not accept the "p value" as low as it may be when you see this in the literature. The correct analysis should include the analysis of variance (ANOVA; one way or factorial) with one of several *post hoc* comparisons, but that is for an article on biostatistics, sorry!

Statistical significance and clinical importance

Significance which is insignificant

RCTs that show a significant difference between the treatments compared are often called "positive". *Statistical significance* is a probabilistic term (the probability to refute a null hypothesis when it is a correct one; the likelihood that the observed difference is, indeed, different from zero).Many clinicians tend to make equivalent a low p (*statistical significance*) with something that is of *clinical importance or significance*. However, "p < 0.0001" has nothing to do with the *magnitude or importance* of a difference or an effect. Such magnitude is called *clinical importance* or significance.

Table 2 - When significance is insignificant

Fact	Consequence	Clinician
Large sample size - huge number of patients	High probability of showing statistical significance	Is the magnitude of the difference clinically important?
$p < 0.05$; *small* difference *in magnitude* (2-3 mg/dL) in 2,000 infants	Statistically different "Bravo!" Practical clinical significance minimal or null	Suspect or doubt the study. Don't change practice yet!
Large number of different studies to show a difference	Real difference is small and probably not important	Suspect or doubt the meta-analysis. Don't change practice yet!

J. Mortimer said: "Lack of belief is an act of faith: The one thing we can be sure of is uncertainty"

The industry advertises a "breakthrough" for hypoglycemia. The authors studied 2,000 hypoglycemic newborns in a randomized, prospective, multicenter, double masked controlled trial (RPMDMCT). One group gets the treatment, the other gets placebo; both receive supplemental glucose. Glycemia increases in the treatment group from a baseline (mean ± SD) of 25±8 mg/dl to 37±6, 46±6 and 52±4 at 30, 60 and 90 minutes. The placebo group goes from 26±9 mg/dl to 35±4 to 43±4 to 50±3. The authors use repeated Student *t* tests (evidently incorrect), reporting a difference in both groups compared to baseline ($p < 0.0001$) and a significantly better response in the treated group compared to placebo at 60 and 90 minutes ($p < 0.001$). Before using this treatment in your patients, and after having read the preceding paragraphs, you would write to the journal and authors asking for ANOVA and exact p values, right? The authors thank you (not very happily though, discussing among themselves who you are to mention their error publicly) and, holding grudges, publish an erratum with the ANOVA. They report $p = 0.048$. Now you feel comfortable using this treatment in your patients. Wait! Before doing so, spend 2 minutes in Table 2 and also ask yourself if the authors measured serum, plasma or whole blood glucose. Did they tell us how they handled the samples and which method they used for actual measurements? With Table 2 and these unanswered questions most sensible clinicians will **not** expose hypoglycemic infants to the new "breakthrough" treatment despite this large, "evident", RPMDMCT. In addition, this treatment may be costly and it may be found later that it produces infrequent, but important adverse effects, not fully analyzed in the study. For "clinical outcomes that matter" and infrequent but important adverse events, keep reading, please!

Nonsignificance which may be significant

By now you know: p value greater than 5% or $p > 0.05$ is "not statistically significant." This means that anywhere between 5.1%-95% of the time the difference is due to chance and the samples are not different. See Table 3 for related concepts and questions a clinician should ask himself or herself in such cases. If still interested in this issue, read later about the right denominators and important outcomes.

Imagine now that the authors of the example above followed 300 of the 2,000 hypoglycemic infants to 5 years of age. By using masked, detailed neurodevelopmental evaluations and careful analyses of potential confounders with logistic regression, they found that some of the developmental and intelligence tests favor the treated group by 7-10 points and that the incidence of cerebral palsy (CP) is half in the treated infants (3% vs. 5.8%). The authors, helped by your previous question and suggestion from a few years before, have now performed excellent statistical analyses. They report no statistical difference, providing the exact p value of 0.059 and an odds ratio (OR) for CP of 0.75 with confidence interval (CI) of 0.67-1.03 which shows no statistical difference, since it crosses 1.0 (Figure 1). Furthermore, they show no adverse effects of the therapy. The question should be: Is this non-statistical significance clinically significant? You'll have to decide! (Table 3). We personally would not give up 7-10 points of our IQ nor would we like to have an almost twofold greater risk for CP. One thing that may be happening here is that there had been no sample size calculations for the effect size on these outcomes and that the sample size at 5 years is not large enough to reach statistical significance for the effect size found (Type II error; see Table 1).

However, as clinicians, we have to decide if the findings are clinically significant and, if so, offer treatment to the patients that entrust their care to us.

Would you agree that, differently from law and justice, a publication is "guilty" until proven innocent? Even the manuscript you are reading now! When assessing and criticizing a scientific publication with healthy skepticism and rational criticism, one is criticizing the publication and NOT the authors. Maybe you have other reasons to criticize the authors, but that doesn't matter in this review.

Denominator

The selection of the denominator is crucial for all studies. Choosing an inadequate denominator takes away the validity

Table 3 - No statistical difference is NOT the same as no clinical significance

A philosophical dilemma is that you can never prove "non-existence" of something
1) Which were the outcome variables studied and how was the variable defined?
2) Who was excluded from the study?
3) Is the "large" sample size the adequate one to use as a denominator for those outcome variables?
4) What is the incidence of the problem in the control group?
5) Could there have been a better sample to study those outcome variables?
6) Would a study with that sample show any differences?
7) When results induced by the intervention are "positive", is that outcome important?
8) If so, which potential adverse events for important outcomes have the authors analyzed?
9) If there are "no significant differences" in poor outcomes, does this mean safety?
10) What is considered "important" by the publication and what do you, your patients and their families consider "important"?
In public health issues, we must be skeptical about whether the absence of evidence of a beneficial result is valid justification for inaction. In public health issues, we must be skeptical about whether the absence of evidence in (infrequent but serious) adverse events is valid justification for action. Be wary of statements like: "This practice has been shown to be safe and should be implemented".

of the results partially or completely, regardless of how prospective, randomized, controlled or blinded the study was and regardless of how fancy the statistics were. Denominators are essential for incidence or rates of diseases, risk factors and impact or effect size of an intervention. The choice of a correct denominator is among the most important factors in the hands of the authors. Of course, the reviewers and editors should be the "first line of protection" when authors choose denominators wrongly or compare some of them erroneously. However, clinicians have the obligation to look carefully at the denominator chosen and at the denominators used in all comparisons and decide what they mean, if anything. Unfortunately, many manuscripts do not make valid comparisons, "flipping" among denominators and/or using incorrect denominators. See the following example.

The rate of prostate cancer decreases over the years in one community as opposed to many other communities in which there has been an increase. Long and behold the denominator used in that community was the entire population (children, young men and women!) Also, during the previous 5 years a proportion of men over 60 moved out of the community because of retirement and weather reasons!

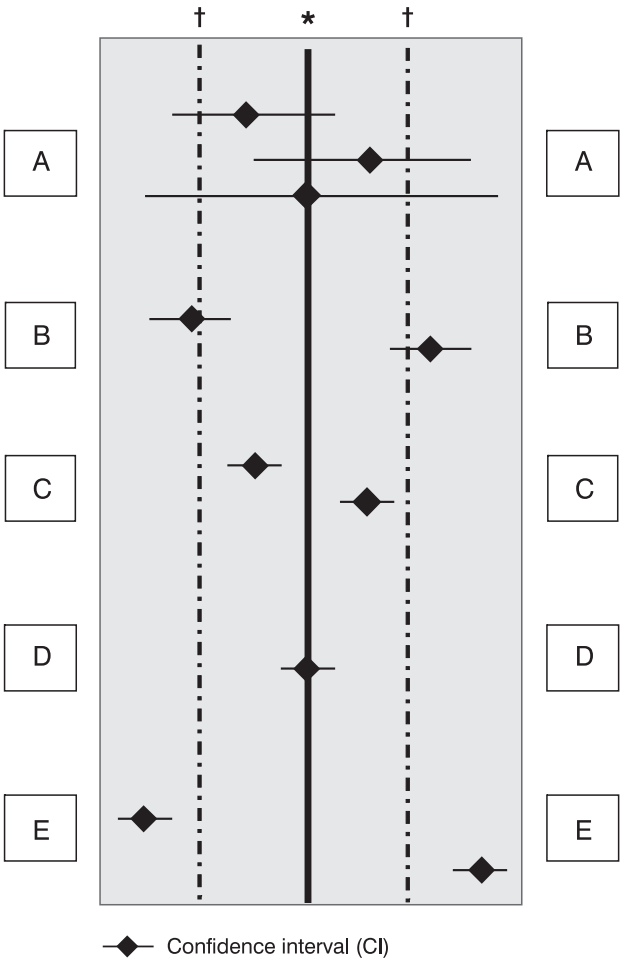
If you look carefully in the manuscripts you read, you will sometimes find significant errors in denominators such as this one (hopefully not as flagrant). Ask yourself: "What is the denominator? What should it be? What is the total population at risk?" In the case of prostate cancer definitely not children, female or young adults. And also be cautious and attentive, since sometimes the authors "change denominators on you". The issue is then to clearly identify who they are talking about and compare this to the real population at risk. For example, in studies of bronchopulmonary dysplasia (BPD) and retinopathy of prematurity (ROP) or severe intraventricular hemorrhage (IVH) do they use all live births < 1.500 g as denominator? Doing so may be similar to the example of prostate cancer above. If many infants die before 6 weeks of age, or are not submitted to detailed ophthalmologic examination or head ultrasounds, the denominator is wrong and rates will be falsely low. When correct denominators are not used, the incidence of the problem is likely to be "better" than it really is! Denominators are of extreme importance when making comparisons or when deciding to change treatments.

Numerators

You also need to identify how the outcome ("numerator") was defined. We hope you agree that it is not the same defining BPD as O_2 for > 28 days as O_2 at home, or need for ventilation for > 4 months.

Number needed to treat (NNT), number needed to harm (NNH), absolute risk reduction (ARR) and relative risk reduction (RRR)

NNT is a popular way of expressing the number of patients who should be treated to prevent one outcome event. It is calculated by obtaining the reciprocal of the ARR (Table 4). For instance, if in a trial of BPD prevention 13% of those treated and 18% controls developed BPD, the ARR would be



*(Solid vertical line) Line of no effect: Relative Risk, Odds Ratio = 1; Risk difference = 0.
† (Dashed vertical lines) Predefined limits of equivalence (threshold for important differences).Adapted from Alderson[70].
A = CI crosses the line of no effect and one or the other or both limits of equivalence: This means that there is insufficient evidence to confirm or exclude an important difference.
B = CI does not cross the line of no effect, but it crosses the limit of equivalence: This means that there is statistical significant difference, but it is uncertain if it is important to patients.
C = CI does not cross the line of no effect and is entirely within the limits of equivalence: This means that there is statistical significant difference, but that it is not important to patients.
D = CI crosses the line of no effect and is entirely within the limits of equivalence: This means no evidence of an important difference. (Remember: it is **not** the same as evidence of no difference)
E = CI is entirely outside the limits of equivalence: This means important difference.

Figure 1 - Relation between confidence interval (CI), line of no effect and threshold for important differences (Predefined limits of equivalence)

5% (18% minus 13%). NNT would be 20 (1/5 x100). Knowing that you have to treat 20 infants to prevent one case of BPD seems more useful than an odds ratio of 0.68 or a RRR of 28% (18% minus 13%/18%).

Thanks to the NNT, you and the parents of your patients may better understand that a treatment may not benefit the individual infant, because for many treatments the NNT is > 10. Clinicians should also wonder about the number needed to harm (NNH) (Table 4). In the ideal world, we would love "certainty", i.e.: NNT of 1, infinite NNH. Since this does not happen, we have to make decisions based on the efficacy of the intervention in terms of risk reduction (Table 4) and possible adverse consequences. If the NNT were relatively low (say 6-10), the most important question to ask is "for what outcome?" If the outcome is important, then an NNT of 10 is always better than an NNT of 100. If you treat 100 infants with an NNT of 10, the *probability of observing no net benefit* is more than 1,000 times lower (i.e.: better) than if the NNT were 100. (Calculated with standard binomial distribution). Don't worry! We will not add formulas for you to learn; we are just trying to clarify a concept. You can choose not to believe this. If you do, please read references in this regard.[3-8] We cannot expand the NNT further, but to say that it focuses only on the treatment arm and neglects the type of placebo or control; it usually overestimates the real value in placebo-controlled trials. Moreover, NNTs are commonly presented as a single value without CI or standard error (Table 4, Figure 1) and this may be insufficient to get "the whole evidence". We suggest you look at an article you have recently read and see if authors report the NNT. If not, by using Table 4, you may be able to do it yourself if data presented in the results clearly describe the incidence in the control and treated groups. If you can do it, you will get a more "evident picture". If you cannot, be wary.

What doesn't work and how to show it

As Alderson wrote,[9] Cochrane posed three key questions about any healthcare intervention: "Can it work?" "Does it work in practice?" and "Is it worth it?" It would be great if the answers were always positive, but in real life, the possible answers might be "yes", "not sure", and "no". The rules for deciding for "yes" are relatively clear and well known, but less has been written about deciding that something doesn't work or does not cause adverse effects, even if infrequent. We will look at issues concerning interventions and dilemmas of trying to decide between an answer of "not sure" and "no," the assessment of what "important outcomes" are and what to do when we are not sure.

Trying to convince the public that a factor has no effect or poses no risk is very hard because this involves "proving a negative".[10] Even with all the "evidence" we can amass, we are many times uncertain about the right treatment choice.[11] It is almost never possible, and in many cases, it is incorrect, to claim that there is no difference in the effects of treatments. One can have truly evidence of absence of an effect or of a difference only if *enough* large and well designed studies were *all* to show that a medical treatment, an exposure or a non-treatment was unassociated with an outcome.[12] One RPMDMCT, as good as it may be, cannot show evidence of absence. In your daily practice you have to consider "other levels of evidence" and not solely RPMDMCT (Table 5). There

Table 4 - Reduction in absolute and relative risks, number needed to treat and confidence intervals

Definitions and concepts	Example 1	Example 2
ICG: verify that it is for the population at risk, with well defined denominator	ICG : 40%	ICG = 10%
ITG = again verify denominators	ITG = 25%	ITG = 5%
RRR = (ICG-ITG)/ICG	RRR= 37.5%	RRR= 50%
ARR = ICG minus ITG ("importance" of the treatment)	ARR = 15%	ARR = 5%
NNT = 1/ARR x 100	NNT = 6	NNT = 20
Is the ICG in the manuscript similar to the population you care for?	If ICG is higher, you will need > NNT in your unit	
How does the ITG compare to your population?	If ITG is similar to or higher than the incidence in your "untreated" population: Don't embark!!	
With the same RRR, if the ICG is low, the NNT will be high		
	ICG = 0.9%	ICG = 90%
	ITG = 0,4%	ITG= 40%
	RRR = 55%	RRR = 55%
	ARR = 0,5%	ARR = 50%
	NNT = 200	NNT = 2
The NNT is not often shown with the corresponding CI. If the ARR for death is 1.3%,the NNT is about 77	If CI for ARR varies between 0.1-2.5%: NNT will vary between 1,000 and 40 (!!!)	

ARR = Absolute Risk Reduction; ICG = Incidence of the problem in the control group; ITG = Incidence of the problem in the treatment group; NNT = Number Needed to Treat; RRR = Relative Risk Reduction.
Number needed to harm (NNH) = An adverse effect due to the exposure that wouldn't have occurred if the treatment had not been used. If for 40 infants treated there is one such effect, the NNH is 40.

will always be uncertainty surrounding estimates of treatment effects; small but important differences can never be excluded in one study.[13] Claims of no effect or no difference may lead clinicians to deny their patients interventions with important beneficial effects or to expose them to interventions with serious harmful effects.[14] Therefore, claims of no effect should be very infrequent, and when they are made we should be skeptical. Reviews or studies making "claims of no effect" or of "evidence of no effect" are erroneous most of the time. Phrases such as "did not reduce", "has no effect" and "is not effective" are usually not justified and should not be allowed by reviewers or editors, because what was shown was "no evidence of effect" as opposed to "evidence of no effect". Acceptable phrases in manuscripts would be: "no significant differences were detected" or "there is insufficient evidence either to support or to refute".

Absence of evidence is no evidence of absence

There is "evident" misconception in these terms, which are not interchangeable. Misinterpretation of "non-significant" findings can become a great problem (Table3). In general, studies with a p value > 0.05 usually can only show *absence of evidence of a difference or absence of evidence of negative effects*. They cannot be considered as "evidence of absence", which *wrongly* implies that the study has shown there is no difference. To interpret these trials as providing evidence of

Table 5 - It is better to obtain an approximate answer to the right question than an exact answer to no question

In your "daily activity"	Suggested "additional levels of evidence'
a) Formulate the question as clearly as possible.There are only answers to formulated questions. A question that has not been asked cannot be answered.	1) One RPMCT is not "the evidence".
b) Seek the evidence (Not just an RCT, not just a conference).	2) Absence of evidence is not evidence of absence.
c) Evaluate the evidence critically.	3) Evidence of no effect is not the same as no evidence of effect.
d) Decide critically if it is applicable in your practice or if it should be eradicated from your practice.	4) Non-randomized clinical trial: there is a *concomitant* control group and uniform evaluation of disease/condition in *both* groups.
e) Evaluate your own results critically.	5) Practical clinical trial.

the ineffectiveness of a treatment or evidence of no adverse events is "clearly wrong and foolhardy".[13] We could use even somewhat harder terms, but we may be accused of being "uncontrolled" and therefore "not evident". It suffices to say that there are dangers in the misinterpretation of nonsignificant results (Table 3). A dramatic example is the fibrinolytic treatment for reinfarction prevention after myocardial infarction (MI), which one of us may need or have already needed. Nineteen of 24 RPMDMCT had "shown no difference" ("p> 0.05") leading to a "statistically significant delay" before the true value of streptokinase, which actually existed in the real world, was appreciated. Our apologies! We got confused and used the term erroneously!! ...The studies did not show "no difference" as we mistakenly wrote; they *only* showed absence of evidence of a difference. Later, a meta-analysis showed a highly significant reduction in mortality (22%). Thank God, serious clinician scientists realized this; some of us will not die due to reinfarction despite "gold standard studies"!

In summary, when issues of public health are of concern, we must be skeptical about whether the absence of evidence is valid justification for action or inaction since where risks are small, or sample size is small, or sample size is large but the correct denominator for the outcome in question is not used in the study, the "negative" p values are likely to be misleading.

In studies of this sort (*wrongly* called "negative studies") wide confidence intervals (CI) many times tell the story or illuminate the absence of evidence.[15] In cases like the streptokinase one, CIs are likely to be wide, indicating "evident uncertainty". Therefore, in cases of absence of evidence in RPMDMCT, we must know the CI and make detailed assessments of important outcomes in that trial. In addition, we need to assess the size of effect and what is important in what situation, observing if authors describe limits of equivalence decided in advance. As shown in Figure 1, if the CI is between those limits of equivalence, an effect is designated as being too small to be important. Please observe Figure 1 where we try to make these concepts clear. Believe us; it is not easy. Just as it is not easy to clearly assess how important a reduction is in the incidence of severe meconium aspiration syndrome (MAS) leading to death or in severe patent ductus arteriosus (PDA) leading to BPD. It is also hard to say who decides when such reductions are important. Of course, if the parents of one unnecessarily seriously affected baby were to decide, they would say that the "difference of 1" was big enough.

Other issues "complicating evidence" is how widespread the exposure is and what the evidence is from previous or subsequent case control or epidemiological studies, which cannot be ignored. Other factors that increase uncertainty of results in RPMDMCT (RCTs, for short) are failure to follow the protocol and non-random loss to follow up. We, as clinicians, are usually not good at understanding all the statistical jargon that has been just mentioned. Authors and journals need to report uncertain results clearly and we should increase our skepticism, trying to incorporate into our daily practices some of the concepts we describe, so we are not misled when the article leaves the impression that it proved that no effect or no difference exists, when evidently it did not.

Important outcomes and outcomes that matter

"The more important things cannot be at the mercy of less important things" (Goethe)

How to define "important outcomes"? Choice of treatment should be determined by effects on outcomes that matter to patients and their parents. Some would even say that they should also matter to society as a whole. Important outcomes are, fortunately, "frequently infrequent", like death, stroke, long-term effects on brain development, vision, hearing, and others. Therefore, to show a statistical difference,

Table 6 - Composite outcomes in neonatology

Compared outcome is *not* a single outcome	Examples/concepts
Strategy for controlling for competing risks or for early effects that may compete with late effects	Having one of the individual outcomes is enough to be in the group with the problem
The main outcome variable is "composed" by two or more outcomes (death, seizures, cerebral palsy)	You know, though, that it is not the same to be dead or to have seizures. (Look at the results of such studies!)
The lower the survival rate, the lower the number of infants with ROP, BPD and adverse outcomes	Prior death and BPD; Prior death and ROP Prior death and adverse neurodevelopmental outcome
Example of "composite outcome" in term infants can be found in Reference 16 on asphyxia and whole body hypothermia. Please, see the individual outcomes included in the composite outcome.	In the "same outcome" the authors include minimal hearing deficit and persistent seizures. These are very different outcomes. An NNT of 6 is reported in the study for the "composite outcome", without much detail, leaving us with uncertainty.

there is a need for a large sample size or a correct denominator of the population really at risk or both. This is why "important" outcomes are commonly poorly studied. So, when reading "evident" manuscripts, try to find out what the authors chose as the main outcome variable. Is it of clinical importance? Is it biologically credible? Are rare and infrequent but serious adverse events (like death) well analyzed and reported?

In summary, be wary of "positive" outcomes of "little" clinical importance, even if $p < 0.0001$, and of "negative" findings of a treatment if nothing is clearly shown about important outcomes (mortality, severe morbidity). The complex issue of "composite outcomes" is in Table 6.[16]

Research to prove is different than to improve: one rct is not evidence

The gold standard for evident-based practice is the RCT. But sometimes gold in some RCTs doesn't shine or is of very low karat. As Jorge Luis Borges said "rational systems extended to the extremes of their rationality turn into nightmares". Of course he wasn't specifically referring to RCTs; as far as we know, none of his stories were randomized.

There are things we know we know and are evident without RCTs. Is it "evident" to you that penicillin cures strep throat? Well, no RCT proves that. Is it evident that if a person jumps from a plane from 1,000 meters and the parachute does not open he will not be healthy after hitting the ground? (No RCT has made this "evident" either). Is it evident to you that the risk of having an accident with harm to oneself or others is higher when driving a car when the brakes do not function or function very inadequately? If you do not know that you know this and would like to do an RCT, be our guest.

RCTs are essential for evaluating the efficacy of clinical interventions if the causal chain between the agent and the outcome is relatively short and simple and where results may be safely extrapolated to other settings. However, it should be no news to anyone that the "first paper is intriguing, with the next three there is growing concern, maybe even a bit of confusion and, after that, what one really would like to know is the real answer." We must be cautious about accepting the results of a single experiment or RCT, by using more extensive exploration under different conditions, in other places and at other times.

However well RCTs might be able to prove or disprove therapeutic claims, however strong their credentials when it comes to seeking evidence, they have their limits when it comes to assuring good care.[17-30] Additionally, RCTs are usually expensive, and always artificial, performed in a selected and restricted group, with exclusion criteria. In brief, RCTs can never be perfect, because they are conducted by humans..., for humans; and they are performed in humans affected by medical conditions and pathologies that are inevitably heterogeneous. RCTs are as good French perfumes, good to smell but not to drink; or like Argentinean wines (Malbec?): great to taste and drink a bit, but not to get intoxicated with.

Some healthcare researchers are advocating to carefully consider "adding karats to the gold standard" with the non-randomized clinical trial (NRCT) and the practical clinical trial (PCT) (Table 5).[19] In RCTs, results are subject to effect modification in different populations. Therefore, both the internal and external validity of RCT findings can be greatly enhanced by observational studies using adequacy or plausibility designs. Additionally, in public health and large-scale interventions, studies with plausibility designs are often the only feasible option and may provide valid evidence of impact when RCTs cannot or are plainly not appropriate.[20] There is also a pressing need for PCTs that are relevant to clinicians and decision-makers. Tunis et al. have addressed very well how to assess their value[21] and Glasgow et al. provided recommendations and examples of how PCTs can be conducted to enhance external validity without sacrificing internal validity.[22]

In summary, developing an evidence base for making public health and practice decisions requires using data from evaluation studies with randomized and nonrandomized designs. Individual studies and studies in quantitative research syntheses require transparent reporting of the study, with sufficient detail and clarity to readily see differences and similarities among studies in the same area. The Consolidated Standards of Reporting Trials (CONSORT) statement provides guidelines for transparent reporting of RCTs. There is also the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND). These guidelines emphasize the reporting of theories used, research design and descriptions of intervention and comparison conditions, and methods of adjusting for possible biases in studies that use nonrandomized designs.[19-22] Many times, NRCT and PCT are far superior to RCTs.

It should be evident to all clinicians that some RCTs don't meet expectations despite a large sample size and that many times they aren't enough to provide "evident evidence" for daily practice. Furthermore, some of them lead to unnecessary confusion and to the production of unnecessary human damage, as infrequent or rare as this may be. We are the ones responsible for the care we deliver, for the care our patients receive, not an RCT or a "superior authority" in a pediatric society. Practical intervention options, alternative research designs and representativeness at the patient level are important to address clinical and policy implications to help reduce the gap between research and practice.

Real life examples

"Evident" RCTs may lead to changes in practice because of hasty analyses and "turning our backs" on previous evidence. We recommend you not skip this section and spend some time on the examples and tables as a form of self reflection, identifying related facts in this summary put forth by many in relation to evidence-based practice. Many show pervasive extremism in neonatology. We, our children, teachers, mentors, friends, dogs and cats and most of our enemies consider that one may gain more insight into life (and for the delivery of care) from "evident" examples, learning from them and from one's own mistakes. But this is not evident since there has been no RCT on the issue! As Popper said: *"If we respect the truth, we must learn from our own mistakes through rational criticism and self criticism"*

Preventing prematurity

This is an extremely important goal! Table 7[23] is used for this example. Remember the analogy with the Malbec wine. Then you decide with "more evidence" what to do as a concerned clinician.

Postnatal steroids

This is a well known, sad story for many babies.[24] Many issues have been unveiled in the last decade. Table 8 shows the risks of generalizing the administration of systemic therapies and miraculously improving short-term effects on one organ in RCTs, but without complete analyses of important outcomes that matter to patient and family.

Carbon dioxide: the good, the bad and the ugly

Table 9 briefly summarizes some of the fascinating stories of neonatal CO_2. Extremism was initially to the low side, then to the high side. A long time ago, based on emerging evidence, we chose to try to prevent both hypocarbia and hypercarbia; it has now been shown that not being CO_2 "extremists" may be good for infants.[44] Table 9 should be fun.

CPAP

Issues are summarized in Table 10. This topic could be by itself in a manuscript. A few references for those interested.[27,28,45-47] Clinical practice is the science of particular or individuals and philosophy is the science of populations' issues. We consider that in practice it is fundamental to evaluate each infant according to his/her needs and not to generalize, even if the concepts seem "very logical" or "evident" or are repeated by "neonatal gurus". But we have not done an RCT.

MAS

Evidently (?), clearing the airway is the first step in resuscitation. Table 11 addresses an RCT which found "absence of evidence" of a beneficial effect of clearing the upper airway in meconium-stained neonates.[48] Before changing practice universally it may be of value to ask some important questions (Table 11). Caution is needed in most (inadequately called) "negative" trials before universalizing practices. As mentioned, the "first paper is intriguing, with the next three there is growing concern, and, after that, what one really would like to know is the real answer." So, as it has been recommended, delivery of care should not change based on only one RCT.[17-22]

Iron and oxidation

Table 12 presents a recent RCT, an editorial and a commentary on that study[29,49,50] in the same volume of a journal in July 2007. Braekke reported (an inadequately called) "negative study". The study does not assess any clinical short- or long-term outcomes, but the effects on clinical outcomes that matter to patients and parents are the effects that should guide practice choices. Basic scientific evidence shows that excess neonatal iron is very damaging[30,51,52] (Table 12). Preventing iron deficiency does not mean that we should use therapies that may induce "iron excess". Evident extremes in neonatology are generally not good (Table 12).

PDA

Affects tiny, preterm infants. Word of mouth and no RCTs "suggest" no intervention. If you decide not to treat, good luck

Table 7 - Decreasing prematurity with 17 α Hydroxyprogesterone (17-OH-P) in a PRDMCT (23)

Questions we must ask	Answer	Comments
Is preventing prematurity an extremely important goal?	Yes	No doubt. However, it is a wide spectrum problem; not all prematurity is the same
How is prematurity defined in the RCT?	< 37 weeks	What are they saying when they conclude?
Is it an important "outcome that matters" to decrease the number of births at 36-37	Maybe	If the costs, adverse outcomes or effects were not high
Who are the authors talking about?	High risk	If you see the ICG, you'll realize how high the risk is.
What do the authors ask? What do they conclude?	Keep reading	If you are very interested, have 23 with you to follow below
What was the incidence of prematurity in the high-risk population *before* starting the study?	36%	This rate seems very, very high. Is this population similar to the one we treat?
What was the incidence of the problem (ICG)?	55%	Much higher than 36% (strange)
What do they say?		Significant beneficial effect to decrease prematurity
What was the rate of prematurity in the treated group (ITG)?	36% (!!)	Any raised eyebrows yet?
What was used as placebo?	Castor oil	Castor oil may induce labor!
Does 17-OH-P prevent prematurity or does placebo (castor oil) increase prematurity?	???	For you to decide
What is the progesterone dose? What was the total number of weekly injections?	???	These and other things are not described in the manuscript!
What is the NNT so that 17-OH-P can prevent prematurity (< 37 weeks)	NNT 6-7	"Looks good"
What is the NNT using the CI reported?	NNT 50-75	For each 50-75 women, one birth <37 weeks would be prevented IN THIS population
What is the NNT to prevent one infant with a birth weight < 1,500 gm?	NNT 250	Up to 250 women would have to be "treated" to prevent one baby < 1,500 gm
How about adverse effects of 17-OH-P?	"No p < 0.05"	One uterine rupture, 5 abortions, 6 fetal deaths, 1 testicular torsion. NND?
Is 17-OH-P safe?	Who knows	Absence of evidence (NOT the contrary)
How effective is 17-OH-P?	You decide	Now as a concerned clinician, you have "more evidence" before you decide
Would you use 17-OH-P for all women at risk?	You decide	You decide with "more evidence" what to do as a concerned clinician before using it

to some of the untreated infants until data on outcomes that matter become available. See Table 13 for related issues.

Cord clamping

A public health issue "affecting" all newborns (see before on RCTs for public health issues). Just for fun, seriously, we "compare" some brief issues of this intervention to no intervention in PDA (Table 13). You have to decide on your own on these two. Stay tuned for rare but important adverse outcomes!

Midazolam

Can you find evident benefits for neonates? Midazolam is "evidently" a neonatal poison, with serious nervous system side effects, more intracranial hemorrhage and death or neurodevelopmental disability, leading neurons to "commit suicide".[54-56] If you continue to administer midazolam, you are probably not among neonatal care providers that try, with all the uncertainty and ignorance we have, to practice based on evidence to improve infants' outcomes, one baby at a time.

Table 8 - Postnatal steroids, lung disease and clinical practice

Issue	Reason
Well known, sad story for many babies[24]	Published RCTs had design flaws or main outcomes of questionable clinical significance
Short term ("unimportant") outcomes	Lung compliance, airway resistance and others
Important outcomes "that matter" not addressed	We do not know why. Do you?
"No problem" with infection, hypertension, hyperglycemia, gastrointestinal bleeding, poor growth, derangements of calcium metabolism and osteopenia	"Absence" of adverse effects, either not reported or the sample size was insufficient to detect differences (Type II errors)
"No problem" with important and long-term outcomes	Not quoted by authors and ignored by clinicians
Previously ignored important concerns regarding the central nervous system[31-33]	Less DNA in cerebrum and cerebellum,[31] decreased head circumference and periventricular leukomalacia in human neonates,[32] among others[33]
Fewer days on intermittent mandatory ventilation (IMV), improved X-Rays, compliance and airway resistance	Beneficial effects of postnatal steroids on lung inflammation, among others
"(In)effective" clinical widespread use of postnatal steroids for prevention or treatment of BPD	All of the above , RCTs[34-36] and recommendations in literature and conferences to practice "evidence-based medicine"
"Since we use postnatal steroids, we do not see BPD any longer" (different ways, doses, periods of time)	Denial; striving for quick fixes; not asking the questions mentioned in this review
What outcomes matter to the patient and family?	Many. Some are more important than others
NND: 7-9 to produce one child with serious CNS sequelae: (CP, neurodevelopment, microcephaly)[25]	Widespread use of postnatal steroids with good intentions and many RCTs
"Low use" of postnatal steroids (2-4% of infants) at University of California San Francisco (UCSF) and Hospital Gonzalez Coro, in Havana	"Evident experience": developed groups who critically reviewed "older" literature and "modern" RCTs. Evident evidence prevented widespread use
Let us not repeat similar mistakes with other therapies	The well being of patients under our care

"All who drink of this treatment recover in a short time, except those whom it does not help, who all die. It is obvious, therefore, that it fails only in incurable cases." (Galen) (not on steroids, of course)

Frequently heard arguments about using or not using some practices are still, unfortunately and evidently, too simplistic. Table 14 shows their possible real meaning with a bit of humor.

Final comments and conclusion

We have summarized important issues related to clinical research findings and their incorporation into our daily encounter with patients. We hope we have shed some light on some significant challenges inherent to health service research. We have correlated the main ideas to many possible clinical applications and used real life examples to emphasize some points. We have provided tools to have a more "evident view of evidence-based practice" and stressed the fact that absence of evidence is not evidence of absence. Misinterpreting a trial that found no significant effect as if "there is no effect" is one of several problems that arise within the more general topic of application of evidence from clinical research in evidence-based care. Most trials in neonatology are far too small to rule out effects of a size that could be clinically important and can also fail to show real evidence of adverse effects. Furthermore, statistical differences may be of a magnitude with no clinical significance or in outcomes that really do not matter much.

It is not easy to find evidence that some therapy or intervention is, indeed, for the better, that it is effective and has (only) the desired effects.[57-70] It is harder to find evidence that some intervention is indeed not necessary in any case, ineffective for all patients, and that not doing it has no undesired important effects, as rare as they may be. One focus of the Cochrane Collaboration Effective Practice and Organization of Care is to include other types of studies beyond RCTs and at to optimize validity, generalizability and evidence of "what works" without causing any unnecessary adverse outcome that matters: To improve professional practice and delivery of effective health services.

Table 9 - Carbon dioxide in neonatal care

Questions to ask	Answer	Comments
Were investigators or care providers blinded to the RCT in 1999[26]?	No	Keep reading.
Was there a clear question?	Yes	"Does it reduce the duration of ventilation?"
What was the sample size?	49	From 114 candidates
Was randomization systematic?	???	65 excluded (24 due to "short ventilation"; 5 "by neonatologist").
What was *not* counted as assisted ventilation?	CPAP	See below
Was CPAP used after extubation?	Yes	According to "clinical indication"
Do these factors suggest potential for systematic bias to you?	You answer	If they do not, it is likely that nothing will
Was the question (hypothesis) answered?	Yes	Authors concluded: the duration of ventilation was not reduced
Was there less BPD or shorter hospital stay?	No	A larger sample size is necessary.
Were the secondary outcomes analyzed in results?	Yes	(No sample size calculation)
Assisted ventilation, respiratory rate, peak inspiratory pressures less in treated group?	Yes	(Do you remember the "criteria" for post-extubation CPAP?)
What was the need for reintubation at < 24 h?	17% treated vs 28% controls	(Do you remember the "criteria" for post-extubation CPAP?)
How important are these two outcomes above to you, the patient and the family?	You decide	(Any potential adverse effects for obtaining such outcomes?)
Any tables with total reintubation rates?	Yes	67% in treated group, vs. 48% in controls
Were there reintubation rates for apnea?	Yes	21% in treated group vs. 12% in controls
Are important undesired effects, like IVH, necrotizing enterocolitis (NEC), ROP and long-term follow up different?	No	Do you know or can you imagine how many babies would be needed to show a difference?
Is absence of evidence evidence of absence?	No	Remember also Type II error
A larger study followed [37]	Yes	Showed no improvement in BPD
We choose not to summarize it due to editorial rules regarding word count	Yes	In life and clinical care, we always have to make choices!
Have you used "permissive hypercarbia", whatever that means?	"Likely"	Not knowing what it means, can we answer yes or no?
When you used it, what outcomes have you measured?	"mmm"	No comments
"Since we use "permissive hypercarbia", we see much less BPD"	"mmm"	No comments
How could you be sure that you were not using "iatrogenic hypercarbia"?	Iatrogenic hypercarbia??	High CO_2 with inadequate alveolar ventilation and more potential for lung injury
What does all available evidence to date say?	"mmm"	Hypercarbia doesn't improve outcomes that matter
What do some selected basic studies say about the bad effects of high CO2?	Many things	High CO_2 and derangements in developing brain and eye [38-40,] and many more
What do recent evident studies show on undesired effects that matter[41,42]?	Nothing good	More need for sedation, more intracranial hemorrhage, worse long-term outcomes
Hypocarbia is bad!	Yes	We've known this for > 25 years[43]
Do we have to be "extremists" and use hypercarbia to prevent hypocarbia?	We did not and do not	It has now been shown that not being "CO_2 extremists" may be good for infants[44]

Somebody once said that in RCTs, randomization should not be left to "random", should not be arbitrary, nonsystematic or left to chance. The same would apply to this practice. (See Galen, Table 8)

Table 10 - CPAP issues (and confusion)

Agree or disagree?	Our answer	Comments
CPAP is an extremely useful and proven tool for neonates	Yes	Ask Gregory et al, 1971
CPAP works for those who need it	Yes	* We hope you agree
CPAP fails or is not enough in severe cases and in many tiny babies	Yes	Even at centers with greatest experience (see the literature)
CPAP is not needed by many babies	True	See*
CPAP is overutilized in many babies who do not need it	True	See*
Myths, surveys, conferences and somewhat "opinionated and religious" beliefs are a plague in CPAP	We think so	See*
The "evidence" (RCTs) is pretty clear with "early prophylactic" CPAP	Yes	To date: No difference and no improvement of important outcomes
The evidence is not as clear with intubation+ surfactant and extubation to CPAP	We think so	To date: No improvement in important outcomes
The evidence about adverse events and risks with early prophylaxis include treatment for babies who do not need it, pneumothorax and ROP	Yes	There are more, even if infrequent: late use of surfactant, flat heads, damaged noses, emergency intubations and CPR, and health-care dollars wasted unnecessarily
CPAP is not a great treatment for pneumothorax, and actually, it increases its rate	Yes	It is "evident". Take a look at the literature and at your own data
Have you heard or said (as with steroids and CO2): "We use CPAP early and a lot, we see much less BPD and *we* have no problems with CPAP"	We heard it over and over	We do not know what neonatologists see or don't see and we are not preoccupied if *neonatologists* have more or fewer problems with CPAP, at least not in this manuscript nor in neonatal clinical care
Use CPAP, of course, but do it with caution	Yes	Use it if indicated, not because "we do so"
One CPAP is the same as any other CPAP	No	Bubble, ventilator, Benviste, Aladdin, etc
Bubble CPAP is "the best"	Not sure	Recent evidence suggests the contrary
We can use CPAP with 100% oxygen	Never	We must blend, humidify and warm gases
Flow and pressures with CPAP are simple	No	No space here to write about it. Be wary!
The best scenario is never to have to use an endotracheal tube, or CPAP, or surfactant	Agree?	We do
With improved prenatal care, there will be no more premature births, no more tubes, CPAP or surfactant	Ha!	Evident "man-made" statement: repeat a statement over and over and it will "become true"
Judicious caregivers would agree with not intubating or using CPAP when not needed and with not using surfactant late when needed	Yes	We must carefully assess each and every infant, individually, and care intensely to decide who needs what.

The question of what good care is cannot be answered with the use of RCTs alone, however impressive their evidence. The answer depends on the character of the desired effect, on what is more important.[18] If there are different goods and bads at stake, a value judgment is called for and this requires careful evaluation, listing and balancing the pros and cons. The task of researchers might still be to provide the evidence that forms the backdrop against which choices may take shape.

In medical practice, we are responsible for each of our patients, not science, RCTs, authors, guideline developers or what an expert says or does. Caregivers have to decide what

Table 11 - A randomized multicenter unmasked trial on meconium-stained amniotic fluid (MSAF)

Questions to ask	Answer	Comments
Is MAS a heterogeneous condition with a wide spectrum of severity?	Yes	Population at high risk for bad outcomes: thick - particulate MSAF
What was the sample size? Where?	Large and multicenter	Keep reading
Who was studied? What was used as denominator?	All those with MSAF	MSAF is thin and watery in about 2/3 of the cases
What was the main outcome variable and how was MAS defined?	Respiratory distress with O2 requirement > 12 hours	Severe MAS is significant respiratory failure leading to IMV/high frequency oscillatory ventilation (HFOV), ECMO and, unfortunately, to some deaths
What was the incidence of MAS as defined that was used for sample size calculation?	7%.	Severe MAS occurs after thick MSAF
What was the study powered for?	20% difference	In MAS as previously defined
What was the study not powered for?	For differences in severe MAS or in mortality	Both are outcomes that matter and occur after thick MSAF
Of the large sample size, how many had thick MSAF?	12%	61% had thin MSAF, which it is "evidently" not associated with severe MAS and death
What was the ICG of MAS as defined?	4 % (not 7% as expected)	Be careful; see sample size calculation
What can happen when the ICG is much smaller than the incidence used for sample size calculation?	Underpowered for the main outcome variable; Type II error	Sample size for a 20% difference with an ICG of 4% can be estimated at close to 4,000
What was the incidence of *severe* MAS in the studied population?	2 %	To show differences, total sample size should be estimated at about 7,000
What was the incidence of mortality in the studied population?	0.6%	To show differences, total sample size should be estimated at > 14,000
What can happen when the incidence of a "secondary outcome" is small?	Underpowered for those outcomes; Type II error	Uncertainty. Is sample insufficient for important outcome "that matters"?
How many deaths in the study?	13	All in the group with thick MSAF
Any differences in mortality between the groups with *thick* MSAF?	Impossible to answer	Sample size is not large enough.
Could the large denominator used have been actually insufficiently large?	Maybe (Type II error)	Maybe there were not enough subjects of real population at risk for bad outcomes
What can be done in cases of wide spectrum and heterogeneity of disease?	Alternative to (potentially small) large sample sizes	Use a sufficient sample size of infants at higher risk as denominator
Should I change practice based on one RCT?	We don't think so	See references in the text.
How safe would it be to do so?	Uncertain	The NND is not known
Is clearing the upper airway not necessary in any case of *thick* MSAF?	You answer	We think we know the answer.
Can you conclude that such practice is ineffective for all infants?	You answer	We think we know the answer.
Has this RCT shown "evidently" absence of undesired, important effects?	You answer	The NND is not known (not enough sample size or incorrect denominator)
Is this study really showing evidence of NO effect?	You answer	Is there simply "no evidence of effect"?
Is this study showing evidence of absence?	You answer	Is it just "absence of evidence"?
Not every baby needs suctioning before delivery of their shoulders, right?	Yes, and differently from …	"Is it not necessary to suction any baby before delivery of the shoulders"
To clear the airway in *all* MSAF infants is an "extreme measure"	True	Not to clear the airway in any infant is another extreme measure
Are neonatologists "extremists" who swing from one extreme to the other?	You decide	When an opportunity is lost, it may be gone forever. When meconium is gone forever into the lungs, the opportunity for prevention is also gone forever.
Is it time to change this?	You decide	
"Can one step into the same river twice?"	You decide	

Table 12 - Iron and oxidation in a randomized trial

Questions to ask	Answer	Comments
Who is being studied?	Healthy fully breastfed preterm infants >6 week old, with vitamin E	No additional oxidant factors
Does the study show *evidence of no effect* of high-dose iron?	No	It shows "no evidence of effect" of high-dose iron on urine oxidative species in a very short period of time (one week only) in a very selected population
Is "no evidence of effect" of high-dose iron on urine oxidative species an "important outcome" that really matters?	No	It does not mean that nothing changed in the present or future intracellular redox status.
Is one week enough time to see changes in the urine concentration of oxidant species?	It may not be	
Can the findings be generalized to all infants under our care?	Not really. Even if the outcomes really mattered	This study says nothing about urine species after iron therapy in younger or sicker infants, infants exposed to other oxidative stresses or on other diets.
Does the study assess oxidation or oxidative injury?	No	See above. Oxidative injury was not studied in this RCT
High-dose iron in infants is well tolerated without indication of increased oxidative injury. Is that true?	Incorrect, incomplete and non-evident.	This statement is made in the commentary to the original article
Are we ready to deliver "evident care" by *early oral administration of* iron to tiny preterm infants, based on three (!!) papers in the same journal?	We are not	If you are, how much will you give and when will you start?
What do related publications say about oxidative species, neonatal iron, oxidative injury and *long-term* cellular toxicity?	It is not the production of oxidative species alone that is responsible for oxidative damage and *long-term* cellular toxicity	The reaction of the species with iron, altering the redox state, may be primarily responsible for the damage and *long-term* toxicity
What do animal and basic studies show?	Neonatal elevation in iron levels produces higher iron content in adult *substantia nigra*, long-term cell loss, enhancement of oxidative injury	Ferrous iron reacts with H_2O_2 producing hydroxyl radicals, damaging proteins, nucleic acids, and membrane phospholipids
Does this study assess clinical short- and long-term effects that matter?	No	It provides important information, but not to change practice universally
What effects on clinical outcomes should guide our practice choices?	Those that matter to patients and parents	When we practice based on those outcomes, we are better physicians
Is preventing iron deficiency extremely important?	Yes. Not doing so leads to abnormal developmental outcomes	Really evident! And it has been so for decades
Does preventing iron deficiency mean that we should use therapy that may induce "iron excess"?	No	Avoiding such excess may be as important as avoiding iron deficiency
Do you remember the three phases of iron metabolism in term and preterm neonates?	You answer, please. Not enough space in this section	If you need help with this and would like us to collaborate, please contact us
How much iron for term babies and when should it be started?	1 mg/kg/d ; after being fully fed	From any source for at least the first year
How much iron should we give preterm babies and from when?	2-6 mg/kg/d. Start no sooner than 4 weeks of age and no later than 8 weeks	From any source, after full feeding; antioxidant mechanisms better developed
What seems judicious?	To avoid iron deficiency and treatments that may cause iron excess	To prevent potential long-term abnormal outcomes of "oxidant extremism"

Table 13 - Issues on PDA and cord clamping

Issue	Late or no treatment for PDA	Late cord clamping
A lot of "noise" about	Not treating a symptomatic PDA in a sick preterm infant.	Clamping the umbilical cord "late" (1-2 minutes -?-) (53-54)
Based on what?	"Word of mouth"/ recent reports (No RCT) Enough "evidence" to change practice?	RCTs. Enough "evidence" to change practice in a public health issue?
For what goal?	Avoid worse effects (not proven) of available treatments vs. no treatment	Prevent iron deficiency anemia (an important goal). Transfusions in preterm babies
Outcomes that matter?	Can't find any carefully analyzed and reported	Only hematological values, ferritin, stored iron, hematocrit and less anemia
Sample size in the "no adverse effects" studies?	Minimal	It varies (40-300 infants), but it may not be enough for rare serious adverse events
Evidence of absence?	No! absence of evidence	No! absence of evidence
Meta-analyses	No	Yes, several
NND for adverse outcomes?	Not carefully assessed	Not carefully assessed
Less important adverse outcomes potentially associated	Days on CPAP, oxygen and furosemide, days on parenteral nutrition and fluids, direct hyperbilirubinemia; osteopenia	Polycythemia, hyperbilirubinemia, respiratory distress, volume and iron load
More important adverse outcomes potentially associated	IMV, severe BPD, weeks of extrauterine malnutrition, head circumference, severe ROP, severe necrotizing enterocolitis	Need for partial exchange transfusion, necrotizing enterocolitis, cerebral stroke and hemorrhage
Previous "evidence"?	Yes, ignored	Yes, ignored
Any concerns?	Maybe? Uncertain? Absence of evidence is not evidence of absence	Maybe? Uncertain? Absence of evidence is not evidence of absence
Question	Are you ready "to wait" in a symptomatic PDA in a sick preterm infant? (Is the baby?)	Are you ready "to wait" to clamp the cord? (Is the baby?)
If so	Carefully quantify and compare rates of all potentially adverse outcomes in the "late treated" infants to infants with no symptomatic PDA or to those treated earlier	Carefully quantify and compare rates of all potentially adverse outcomes in the "late treated" infants to other infants
Suggestion/reflection	Persistently prolonged PDA ("PPPDA") may be associated with persistently prolonged pulmonary and developmental abnormalities (a different but important "PPPDA")	Are there other ways of preventing iron deficiency and poor long-term outcome? Are they associated with less potential damage for "innocent bystanders"?

You will have to decide on your own on these two, just as we had to do. Stay tuned!Good luck to untreated or late treated infants until data on outcomes that matter (even if rare) become availableMany times, the problems or a significant problem in an infant are due to the proposed solutions

to do using evidence, including the uncertainties and definitely the important outcomes that really matter to patients, family and society. What makes each of us more or less responsible in the care we deliver is not only what we decide and/or accept to do, but also what we refuse to do. It is our responsibility as clinicians caring for patients to practice according to this understanding.

It is necessary to create a culture that is comfortable with estimating and discussing uncertainty. We are hopeful that in

Table 14 - Arguments used to justify various practices and their real meaning

Argument	Possible real meaning
1) "In my experience this works"	Successive repetition of mistakes
2) "We do not see any problems with this procedure or treatment"	Not looking correctly at their available data. (Rare, albeit important, events are hard to quantify)
3) "I have done this over and over and in 'case after case' with great results"	Saw two cases, maybe three, and/or selective observation capabilities. (Denial is frequent)
4) "It works well for *us*"	Nobody cares how it works for care providers ("us")
5) "*We* do not have any problems with this procedure or treatment"	Again, one only cares if the patients have the problem (not the care providers)
6) "I have never seen that problem"	Either he/she does not work, does not pay attention, does not treat infants at risk or is on vacation when the problems occur
7) "We do it because so and so (usually a "recognized" neonatologist) says that"	I cannot think by myself and I follow opinions and conferences like "pulpit dictums and sermons"
8) "We do it because that is what they do at such and such (prestigious) place or university"	So what? Many things that are not correct are done by a lot of people, including chairmen and others
9) "This is the way we do it here, have done it for over 10 years and we have had good results"	Don't confuse me with the facts, my mind is already made. (Additionally, does what we say we do happen to all patients?)
10) "We don't do that here"	See above.
11) "There is not enough evidence and more RCTs are needed"	This sounds impressive (More frequently than not doesn't know what evident evidence is and has heard the point at some conference. See argument 12 also)
12) "The studies are not clear"	Has only read abstracts or manuscripts superficially (usually from 10-15 years ago. See argument 11 also)
13) "Where I worked before they did it that way"	So what? See 8-10 above for the real meaning
14) "Where is the evidence?"	What is evidence for those who use this argument simplistically? How much do they engage in critical discussions?
15) "He/she does it that way, and some others change that when they come on call or on service"	Anarchy. (Is what he/she does more important than what the patient needs?)
16) "We will defer to the opinion of our expert in the topic"	See argument 9. And what is an expert for you? (The opinion of experts is not the same as expert opinion)
17) "We try hard to ensure that the patients under our care receive best available and effective care"	Good! (see the "evidence" critically, analyze your data, quantify adverse events, try to prevent damage)

the future changes in practice will occur with studies using the correct denominators and accuracy in interpretation and language. With this and with the incorporation of *all* available research (not just only one or the latest RCT) uncertainty can be reduced. As this happens more and more, fewer guidelines that are developed and implemented "universally" will

be proven at a later time to be wrong, thereby increasing important outcomes that matter and the well being of more infants.

We hope that, after reading (and re-reading) this manuscript, you'll feel more empowered to make the right choices for your patients and that it becomes evident that some expert opinions (and not opinion of experts) can become yours too. To end, our best wishes for the delivery of care you provide your patients with, one baby at a time, in this complex era of "evidence-based practice".

Acknowledgements with evident gratitude

Drs B. Griffin and F. Bednarek, University of Massachusetts, Mentors "par excellence", positively influenced our approach to caring intensely

Drs R. Phibbs, G. Gregory and J.A. Kitterman, Neonatology, CVRI at University of California, San Francisco, Mentoring, teaching and caring in daily practice positively influenced our approach to practice.

Drs J. Comroe, J. Clements, A. Rudolph and S. Glantz, CVRI at University of California, San Francisco, Contributed to our education on care and statistics through their advice, lectures and publications.

Dr E. Dueñas, Havana, Cuba, Inspirational in organization of care and outcome improvement.

Dr J. Sinclair, McMaster University, Canada, Workshops, lectures, publications. Inspirational.

Dr A. Brann, Emory University, Untiring clinician and information seeker.

Drs G.R. Norman and D. Streiner, Canada, Workshop and book Biostatistics: the bare essentials.

Dr P. Alderson, UK Cochrane Centre, His writings as model.

Others, Many, Writings, lectures, approach to practice, friendship.

If you have read slowly and attentively, it should all be evident!

References

1. Norman GR, Streiner DL. Biostatistics: the bare essentials. 2nd ed. Hamilton: B.C. Decker; 2000.
2. Michael M, Boyce TW, Wilcox AJ. Biomedical bestiary: an epidemiologic guide to flaws and fallacies in the medical literature. Boston: Little, Brown;1984.
3. Alderson P , Deeks JJ. Concerning: 'The number needed to treat needs an associated odds estimation'. J Public Health (Oxf). 2004;26:400. Author reply in: J Public Health (Oxf). 2004; 26:401.
4. Aino H, Yanagisawa S, Kamae I.The number needed to treat needs an associated odds estimation. J Public Health (Oxf). 2004;26:84-7. Erratum in: J Public Health (Oxf). 2004;26:218.
5. Artalejo FR, Banegas JR, Artalejo AR, Guallar-Castillo, P. Number-needed-to treat to prevent one death. Lancet. 1998;351:1365.
6. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ. 1995;310: 452-4.
7. Saver JL. Deriving number-needed-to-treat and number-needed-to-harm from the Saint I trial results. Stroke. 2007; 38: 257.
8. de Craen AJ, Vickers AJ, Tijssen JG, Kleijnen J. Number-needed-to-treat and placebo-controlled trials. Lancet. 1998;351: 310.
9. Alderson P, Groves T. What doesn't work and how to show it. BMJ. 2004;328: 473.
10. Hunt A, Millar R. AS science for public understanding. Oxford: Heinemann, 2000.
11. Chalmers I. Well informed uncertainties about the effects of treatments. BMJ. 2004;328: 475-6.
12. Joffe M. "Evidence of absence" can be important. BMJ. 2003; 326: 1267.
13. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ. 1995;311: 485.
14. Alderson P, Chalmers I. Survey of claims of no effect in abstracts of Cochrane reviews. BMJ. 2003;326: 475.
15. Altman DG, Bland JM: Confidence intervals illuminate the absence of evidence. BMJ. 2004;328:1016-7.
16. Shankaran S, Laptook AR, Ehrenkranz RA, Tyson JE, McDonald SA, Donovan EF, et al. Whole-body hypothermia for neonates with hypoxic-ischemic encephalopathy.N Engl J Med. 2005; 353:1574-84.
17. Sinclair JC. Weighing risks and benefits in treating the individual patient. Clin Perinatol. 2003;30: 251-68.
18. Mol A. Proving or improving: on health care research as a form of self-reflection. Qual Health Res. 2006;16:405-14.
19. Des Jarlais DC, Lyles C, Crepaz N; TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94:361-6.
20. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. Am J Public Health. 2004; 94: 400-5.
21. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA. 2003;290: 1624-32.
22. Glasgow RE, Magid DJ, Beck A, Ritzwoller D, Estabrooks PA. Practical clinical trials for translating research to practice: design and measurement recommendations. Med Care. 2005; 43: 551-7.
23. Meis PJ, Klebanoff M, Thom E, Dombrowski MP, Sibai B, Moawad AH, et al. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. N Engl J Med. 2003; 348: 2379-85.
24. Finer NN, Craft A, Vaucher YE, Clark RH, Sola A. Postnatal steroids: short-term gain, long-term pain? J Pediatr. 2000; 137: 9-13.
25. Mariani G, Cifuentes J, Carlo WA. Randomized trial of permissive hypercapnia in preterm infants. Pediatrics. 1999;104: 1082-8.
26. Carlo WA, Stark AR, Wright LL, Tyson JE, Papile LA, Shankaran S, et al. Minimal ventilation to prevent bronchopulmonary dysplasia in extremely-low-birth-weight infants. J Pediatr. 2002;141: 370-4.

27. Verder H, Robertson B, Greisen G, Ebbesen F, Albertsen P, Lundstrom K, et al. Surfactant therapy and nasal continuous positive airway pressure for newborns with respiratory distress syndrome. Danish-Swedish Multicenter Study Group. N Engl J Med. 1994;331: 1051-5.

28. Vain NE, Szyld EG, Prudent LM, Wiswell TE, Aguilar AM, Vivas NI. Oropharyngeal and nasopharyngeal suctioning of meconium-stained neonates before delivery of their shoulders: multicentre, randomised controlled trial. Lancet. 2004; 364: 597-602.

29. Braekke K, Bechensteen AG, Halvorsen BL, Blomhoff R, Haaland K, Staff AC. Oxidative stress markers and antioxidant status after oral iron supplementation to very low birth weight infants. J Pediatr. 2007;151: 23-8.

30. Hutton EK, Hassan ES. Late vs early clamping of the umbilical cord in full-term neonates: systematic review and meta-analysis of controlled trials. JAMA. 2007;297: 1241-52.

31. Howard E. Reductions in size and total DNA of cerebrum and cerebellum in adult mice after corticosterone treatment in infancy. Exp Neurol. 1968;22:191-208.

32. Regev R, de Vries LS, Noble-Jamieson CM, Silverman M. Dexamethasone and increased intracranial echogenicity. Lancet. 1987;1: 632-3.

33. Noble-Jamieson CM, Regev R, Silverman M. Dexamethasone in neonatal chronic lung disease: pulmonary effects and intracranial complications. Eur J Pediatr. 1989;148: 365-7.

34. Stark AR, Carlo WA, Tyson JE, Papile LA, Wright LL, Shankaran S, et al. Adverse effects of early dexamethasone in extremely-low-birth-weight infants. National Institute of Child Health and Human Development Neonatal Research Network. N Engl J Med. 2001;344:95-101.

35. Murphy BP, Inder TE, Huppi PS, Warfield S, Zientara GP, Kikinis R, et al. Impaired cerebral cortical gray matter growth after treatment with dexamethasone for neonatal chronic lung disease. Pediatrics. 2001;107: 217-21.

36. Barrington KJ. The adverse neuro-developmental effects of postnatal steroids in the preterm infant: a systematic review of RCTs. BMC Pediatr. 2001;1: 1.

37. Gu XQ, Xue J, Haddad GG. Effect of chronically elevated CO2 on CA1 neuronal excitability. Am J Physiol Cell Physiol. 2004; 287:C691-7.

38. Martin C, Jones M, Martindale J, Mayhew J. Haemodynamic and neural responses to hypercapnia in the awake rat. Eur J Neurosci. 2006;24: 2601-10.

39. Fritz KI, Zubrow A, Mishra OP, Delivoria-Papadopoulos M. Hypercapnia-induced modifications of neuronal function in the cerebral cortex of newborn piglets. Pediatr Res. 2005;57: 299-304.

40. Kaiser JR, Gauss CH, Pont MM, Williams DK. Hypercapnia during the first 3 days of life is associated with severe intraventricular hemorrhage in very low birth weight infants. J Perinatol. 2006; 26: 279-85.

41. Thome UH, Carroll W, Wu TJ, Johnson RB, Roane C, Young D, et al. Outcome of extremely preterm infants randomized at birth to different PaCO2 targets during the first seven days of life. Biol Neonate. 2006;90: 218-25.

42. Sola A, Spitzer A, Morin FC 3rd, Schlueter M, Phibbs RH. Effects of arterial carbon dioxide tension on newborn lamb's cardiovascular response to rapid hemorrhage. Pediatr Res. 1983;17:70-6.

43. Fabres J, Carlo WA, Phillips V, Howard G, Ambalavanan N. Both extremes of arterial carbon dioxide pressure and the magnitude of fluctuations in arterial carbon dioxide pressure are associated with severe intraventricular hemorrhage in preterm infants. Pediatrics. 2007;119: 299-305.

44. Gregory GA, Kitterman JA, Phibbs RH, Tooley WH, Hamilton WK. Treatment of the idiopathic respiratory-distress syndrome with continuous positive airway pressure. N Engl J Med. 1971; 284: 1333-40.

45. Finer NN, Carlo WA, Duara S, Fanaroff AA, Donovan EF, Wright LL, et al . Delivery room continuous positive airway pressure/ positive end-expiratory pressure in extremely low birth weight infants: a feasibility trial. Pediatrics. 2004;114:651-7.

46. Sandri F, Ancora G, Lanzoni A, Tagliabue P, Colnaghi M, Ventura ML, et al. Prophylactic nasal continuous positive airways pressure in newborns of 28-31 weeks gestation: multicentre randomised controlled clinical trial. Arch Dis Child Fetal Neonatal Ed. 2004;89:F394-8.

47. Morley C; COIN Trial Investigators. COIN Trial Abstract at PAS Meeting, Toronto, Canada, May 2007.

48. Jobe AH. Iron for preterm infants. J Pediatr. 2007;151:A3.

49. Kling, P J. Iron supplementation in prematurity: how much is too much? J Pediatr. 2007;151:3-4.

50. Zecca L, Youdim MB, Riederer P, Connor JR, Crichton RR. Iron, brain ageing and neurodegenerative disorders. Nat Rev Neurosci. 2004;5: 863-73.

51. Kaur D, Peng J, Chinta SJ, Rajagopalan S, Di Monte DA, Cherny RA, Andersen JK Increased murine neonatal iron intake results in Parkinson-like neurodegeneration with age. Neurobiol Aging. 2007;28:907-13.

52. Peng J, Peng Li , Stevenson FF, Doctrow SR, Andersen JK. Iron and paraquat as synergistic environmental risk factors in sporadic Parkinson's disease accelerate Age-related neurodegeneration. J Neurosci. 2007;27:6914–92.

53. Chaparro CM, Neufeld LM, Tena Alavez G, Eguia-Líz Cedillo R, Dewey KG. Effect of timing of umbilical cord clamping on iron status in Mexican infants: a randomised controlled trial. Lancet. 2006;367:1997-2004.

54. Anand KJ, Barton BA, McIntosh N, Lagercrantz H, Pelausa E, Young TE, et al. Analgesia and sedation in preterm neonates who require ventilatory support: results from the NOPAIN trial. Neonatal Outcome and Prolonged Analgesia in Neonates. Arch Ped Adol Medicine. 1999;153:331-8.

55. Waisman D, Weintraub Z, Rotschild A, Bental Y Myoclonic movements in very low birth weight premature infants associated with midazolam intravenous bolus administration. Pediatrics. 1999;104:579.

56. Olney JW, Young C, Wozniak DF, Jevtovic-Todorovic V, Ikonomidou C. Do pediatric drugs cause developing neurons to commit suicide? Trends Pharmacol Sci. 2004;25: 135-9.

57. Sinclair JC. Evidence-based therapy in neonatology: distilling the evidence and applying it in practice. Acta Paediatr. 2004; 93: 1146-52.

58. Sinclair JC, Haughton DE, Bracken MB, Horbar JD, Soll RF. Cochrane neonatal systematic reviews: a survey of the evidence for neonatal therapies. Clin Perinatol. 2003;30: 285-304.

59. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. 1996. Clin Orthop Relat Res. 2007;455: 3-5.

60. Sacket DL. Patients and therapies: getting the two together. N Engl J Med. 1978;298: 278-9.

61. Sackett DL. Superiority trials, noninferiority trials, and prisoners of the 2-sided null hypothesis. ACP J Club. 2004;140:A11.

62. Chalmers I. Proposal to outlaw the term "negative trial". BMJ. 1985:290:1002.

63. Chalmers I. Well informed uncertainties about the effects of treatment. BMJ. 2004:328:475-6.

64. Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 15. Disseminating and implementing guidelines. Health Res Policy Syst. 2006;4:27.

65. Mainland D. Is the difference statistically significant? An innocent request for quack treatment. Clin Pharmacol Ther. 1969; 10: 436-8.

66. Mainland D. Statistical ward rounds --14. "How many patients would we need?Clin Pharmacol Ther. 1969;10:272-81.

67. Mainland D. Medical statistics-thinking vs arithmetic. J Chronic Dis. 1982;35: 413-7.

68. Mainland D. Statistical ritual in clinical journals: is there a cure?--II. Br Med J (Clin Res Ed). 1984;288:920-2.

69. Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al. Need for expertise based randomised controlled trials. BMJ. 2005;330: 88.

70. Alderson P. Absence of evidence is not evidence of absence. BMJ. 2004;328: 476-7.

Correspondence:
Augusto Sola MD
MANA and Atlantic Neonatal Research Institute, Division of Neonatology, Atlantic Health System,
Morristown, NJ
100 Madison Ave
Morristown, NJ 07960
Tel.: (+1) 973-971-8985
E-mail: augustosolaneo@gmail.com