Dear readers and authors!

We are grateful to all of you for participating in our journal. Authors — for your work, readers — for reading. In fact, it's hard to say which of these is more important and which motivates all of us more. Analyzing the articles published in our journal last year, attention is drawn to the sometimes different approach of the authors to the methodology, assessment of the results, sample size, patients lost to follow-up, etc. Evaluation of research results is always based on three points:

1) Whom do we evaluate?
2) What and how do we evaluate?
3) How long do we follow-up?

Now 2020, the era of big-data and artificial intelligence, but these "commonplace" questions are still relevant. On the other hand, we are not unique. In the 2000s and the first half of the 2010s, foreign top-rated journals published a series of editorials that dealt with the typical defects in submitted and published articles. We also have to go this way, albeit with a 10-year delay.

Whom do we evaluate? This issue includes internal validity (inclusion and exclusion criteria), sample size, and patients lost to follow-up rates. Let us dwell on the latter in more detail. The patients loss to follow-up in the course of research is common. Incorrect losses, which reduce the value of the study, can be talked about in two cases: when the percentage of losses from the compared groups is very different, or when the lost patients differ from those who remain. Among the lost patients there may be not only those who are asymptomatic, completely satisfied and simply do not want to remember and return, but also, on the contrary, absolutely disappointed. And we never know what patients in the lost group are actually more.

What percentage of losses can be regarded as acceptable? There is no clear answer to this question [1]. But more often we use the rule that dropping out <5% leads to little bias, and dropping out > 20% means a serious threat to validity [2]. In any case, it should be understood that even a small loss rate could seriously limit the study [3], and sometimes loss <20% can be critical. There is a methodological rule to suspect the worst scenario in every case that is lost [2].

Speaking about the number of patients included in the study, in the overwhelming majority of cases, there is a passive approach in our works: we include as many patients as there were at the time the authors decided to draw the line and start the analysis. As a result, the reliable differences are obtained for some of the compared parameters, and unreliable — for others. But are they really unreliable (type II error, β-error), or will the reliability restore with an increase in the number of patients? This can be avoided with simple online sample size calculators. But in papers submitted to the journal, authors rarely ask the question: "Have they included enough patients in their study and are they publishing it too early?"

What and how do we evaluate? This issue includes the problem of determining the end points, the validity of the scales used, etc. An endpoint is defined as an outcome where participants are typically excluded from further research. A "surrogate" endpoint is a criterion that is not itself the desired outcome of the treatment. The "surrogate" endpoint may be the normalization of some laboratory blood test, which in fact may not be of particular importance for patients, although it may be a proven risk factor for something.

It is important not to make a mistake with the choice of endpoints. For example, the authors decided to study the results of diaphyseal fractures treatment and chose the classic endpoints — the rate of union/non-union fractures and range of motion in the adjacent joints, and the results were evaluated within six months or a year after the surgery. Such an approach would have been understandable to us 40 to 70 years ago, but did the authors correctly identify the end point in their study? What does this work give us in 2020? In addition to these endpoints, the diaphyseal fractures have another important endpoint — arthrosis of the adjacent joints. Of course, to evaluate it, a much longer follow-up is needed. But what prevents the inclusion in the work a "surrogate" criterion of mechanical axes deviation? What are the limits of acceptable deviations? Of course, the larger the displacement, the worse, but we cannot say that we know everything in this issue. If such a work included a "surrogate" criterion of axial deviation, residual displacement, and a public commitment was given on the journal pages to trace the real development of arthrosis in 5 to 7 years, and even in 2 to 3 years, then such work would be much more valuable, and the labor costs of the authors themselves for such a series of works were hardly large.

Here's one more example of the endpoint selection problem. In evaluation of oral anticoagulants effectiveness as an endpoint, it seems logical to choose a thromboembolic event. And this has been the case

for almost 15 years. But back in the late 2010s, some orthopedists noted an increase in infectious complications after arthroplasty against the background of oral anticoagulants. Although, the reliability of an increase in the rate of infections after arthroplasty has not been proven due to the single-center nature of the studies and the low incidence of infections in general. It is clear that this problem could be solved only within the framework of a very large multicenter study, which was started only in 2016, 10 years after the beginning of the era of oral anticoagulants, and will be completed in 2021 (PEPPER Trial, 25 thousand patients [4]).

The problem of endpoints choosing is actually not that simple, not new, still relevant, and there is no shame in its recognition. We are not unique in the problem of endpoints choosing and we follow the same path as our foreign colleagues. In the AAOS 2012 Position Statement was noted: "The AAOS believes that systems for measuring and reporting quality in health care should continue to evolve and expand. The current generation of quality measures, which primarily rely on process measures and administrative data, have not yet been proven to accurately correlate with improved functional outcomes, which are the primary outcomes of interest to patients who undergo orthopaedic procedures."[5]. For example, almost all arthroplasty registries are implant-oriented, not patient-oriented. They do not allow assessing the functional outcome. And this paradigm must change [6]. Of course, this cannot happen quickly.

Let's say a banal thing — scales are used to evaluate functional results. Some hospitals use the scales for all patients, whether they are included in a study or not. Such an approach sometimes irritates practitioners, for example, the American orthopedists note that it is "difficult, cumbersome, time consuming and expensive" [7]. Yes, this direction is stalled in our country too. But we believe that the medical charts digitalization will someday become "friendly", and this issue of routine scale assessment will be resolved.

Here it is important to pay attention to what scale(s) we are going to use. It is very gratifying to see more and more often in the pages of our journal some works devoted to cross-cultural adaptation of scales. The work on cross-cultural adaptation of scales is undoubtedly important, but it is not always clear why the authors chose this particular scale. It would be desirable for the domestic orthopedic community to come to a common understanding what scales should be used. One way or another, we will come to a uniform digital contour of medical documentation. And for the "compatibility" of the data evaluated in the future, this work on the unification of the scales should be supported and continued in every possible way.

How long do we follow-up? Above, this issue was already touched when we talked about evaluating the results of diaphyseal fractures treatment. It is completely unacceptable when the authors choose an intolerably short observation period, despite the fact that other works on the similar problem have already pointed out the importance of the long-term consequences. On the other hand, the absence of other work on the study of long-term results does not mean that the issue under study is only of short-term relevance. For example, after the first cemented vertebroplasty performed by P. Galibert and colleagues in France in 1987 [8], enthusiasm arose, the indications for vertebroplasty were significantly expanded, and for many years the authors published impressive positive results with a follow-up of 1 to 2 years. However, more long-term results have revealed serious negative consequences of this technology and significantly reduced the scope of its application.

The issue of continuity in evaluation the results of orthopedic pathology treatment is very important. Formally, in our country, the "children's" and "adult" orthopedics are one specialty. In fact, only in a number of institutions do they exist together, and most of them live a separate life. The long-term results of the treatment of children with orthopedic pathology are, at best, limited to 18 years. This is absolutely insufficient to understand the effectiveness of the methods of treatment used for a adeqately large group of diseases. The imminent digitalization of medical records can help with this. And after years or decades, we will finally be able to access children's X-ray images of an adult patient always, not only in cases where the child has a gross orthopedic problem and the entire history of the images was carefully kept by the parents. For example, it could be recalled the problem of undiagnosed small forms of epiphysiolysis of the femoral head and its possible role in the development of CAM-impingement in decades.

It is important that researchers always clearly understand how the issues they study relate to the duration of their observation. In this matter, excessive criticality, which will be refuted by time, is more useful than premature confidence.

How to interpret the results? In this matter, we want to pay attention to the statistical and clinical significance, which, unfortunately, are not always understood by the authors. If statistical significance is obtained, then before making a conclusion, you need to ask yourself the question: was the true endpoint estimated, or were significant differences obtained by a surrogate criterion? For example, the authors studied the rate of dislocations after arthroplasty with different approaches and found that with one approach, the cup anteversion was significantly different from anteversion with the other approach. In this case, the end point is the dislocation, not ante-

version, although it (or rather its absence) is considered a reliable risk factor for dislocation. The conclusion that the risk of dislocation increases in a group with such and such access would be false. Or another example — asymptomatic small vein thrombosis of the leg veins after surgery, which is actually not so important for the patient, and attempts to correct them lead to an increase in the same hemorrhagic complications. Unfortunately, such a banal error of conclusion based on the "surrogate" criterion in one or another variation is not uncommon.

Another important point — are the reliable benefits obtained as a result of the study important for a particular patient? This recalls a large study of several thousand patients in which the use of probiotics for the gastrointestinal symptoms treatment after antibiotic therapy was studied. As a result, it turned out that the duration of symptoms in the probiotic group was significantly shorter by 0.4 days. How important was it to the patient really? The authors of that study thought it was important.

The question is natural: what follows from the examples given, and what will be the wishes for the further work? It should be noted that the era of local study of several case histories in the archive is confidently receding into the past. Such an approach remains relevant only for very narrow issues. The solution is in registers and digitalization. They will make it possible to more reliably evaluate the results of treatment and identify their weaknesses. At the same time, on the path of "registration" and digitalization, we need to solve the problem of functional results evaluation and important results for the patient in general. Formulating conclusions, each researcher should ask himself the question — how clinically significant are the results obtained, whether this effect was really worth the cost, and whether the clinical conclusion is based on a "surrogate endpoint"? In addition, authors should be aware that the results obtained are the results of today, and they may change. Planning a study, it is always important to ask the question: is the follow-up period sufficient for the particular parameter being studied? On the other hand, one should not be afraid to present the results as preliminary with the obligation to study more long-term outcomes.

Литература

1. Sacket D.L., Richardson W.S., Rosenberg W. Evidence-Based Medicine: How to Practice and Teach EBM. New York: Churchill Livingstone; 2000. 261 p.
2. Dettori J.R. Loss to follow-up. *Evid Based Spine Care J.* 2011 Feb; 2(1): 7–10. doi: 10.1055/s-0030-1267080.
3. Bhandari M., Guyatt G.H., Swiontkowski M.F. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. *J Bone Joint Surg Am.* 2001;83(6):916–926. doi: 10.2106/00004623-200106000-00015.
4. Comparative Effectiveness of Pulmonary Embolism Prevention After Hip and Knee Replacement (PEPPER). https://clinicaltrials.gov/ct2/show/NCT02810704
5. American Academy of Orthopaedic Surgeons. Position statement. Public reporting of provider performance. https://www.aaos.org/globalassets/about/position-statements/1183-public-reporting-of-provider-performance.pdf .
6. Franklin P.D., Allison J.J., Ayers D.C. Beyond joint implant registries: a patient-centered research consortium for comparative effectiveness in total joint replacement. *JAMA.* 2012;308:1217-1218. doi: 10.1001/jama.2012.12568.
7. Ayers D.C., Bozic K.J. The Importance of Outcome Measurement in Orthopaedics. *Clin Orthop Relat Res.* 2013; 471(11):3409-3411. doi: 10.1007/s11999-013-3224-z.
8. Galibert P., Deramond H., Rosat P., Le Gars D. [Preliminary note on the treatment of vertebral angioma by percutaneous acrylic vertebroplasty]. Neurochirurgie. 1987;33(2):166-168. (In French).