7–10

Heterogeneity of treatment effects: from "How to treat" to "Whom to treat"

Authors Joseph R Dettori[1], Daniel C Norvell1, Andrea C Skelly[1], Jens Chapman[2]
Institutions [1]Spectrum Research Inc, Tacoma, WA, USA
[2] Harborview Medical Center, Seattle, WA, USA

The classic approach in clinical medicine and medical research has been to evaluate how to treat a specific illness or disease using various study designs to compare treatments and possibly sophisticated statistical analyses to explain effects that take into account confounding factors. Examples from infectious disease and trauma research provide excellent examples of how this approach has been successful in delineating the best options for clinical decision making and treatment.

This classic approach may be incomplete, however, in the evaluation of how to treat complex and/or chronic conditions. For such conditions, even randomized controlled trials (RCTs) and other methodologically rigorous studies may show only a small-effect size or fail to find significant differences in outcomes between treatment groups. A number of factors, which are beyond the scope of this article, may contribute to such findings. The bottom line is that we are frequently left with inconclusive evidence about how to treat a given disease. Treatment studies of chronic low back pain and many forms of cancer are examples of entities for which clinical decision making is complex and evidence on the best treatment may be inconclusive.

This conundrum invites us to expand our thinking and consider new approaches to both research and clinical management. One newer conceptual approach is the "targeted treatment approach" wherein the ideal treatment by the ideal provider is paired with the ideal patient. This represents at the very least a partial paradigm shift from "How to treat" to "Whom to treat."

How do we get to "Whom to treat?"

Although the concept of exploring heterogeneity is routinely applied to epidemiological studies, concentrated application of it to clinical research in spine disease has not been done frequently. The basic concepts related to evaluation of heterogeneity are presented below.

A clinical trial seeks to answer the question, is treatment A better than treatment B *on average for a select population?* However, clinicians seek an answer to a different question: Is treatment A better than treatment B *for this specific patient?* The best treatment for a population may not be the same as for the individual patient. Why is this so? Because the same treatment often produces different results in different patients; some receive substantial benefits, many receive little or no benefit, and a few are harmed [1]. The term that describes this situation is "heterogeneity of treatment effect" (HTE).

What factors influence HTE?

Kravitz et al [1] describes four factors that influence HTE:

- Risk of disease without treatment: The risk of disease progression without treatment represents the prognosis of the patient who receives no treatment, a placebo treatment, or a standard (nonexperimental) treatment. It is similar to the natural history of the disease.

- Responsiveness to treatment: The responsiveness to treatment relates to the probability that a patient will experience a benefit from the treatment. This can depend on nonpatient-related factors (eg, the technique of the surgeon, the effectiveness of an implant, or the concentration of a biologic at the target site) or on characteristics of the patient (eg, comorbidities or genetic differences).

- Vulnerability to adverse events: Vulnerability to adverse events is the likelihood that a patient will experience a side-effect that would not occur in the absence of the treatment. Whether a patient experiences treatment-related or disease-related events will often depend on the context. For example, adjacent segment disease following fusion may be either treatment or disease related.

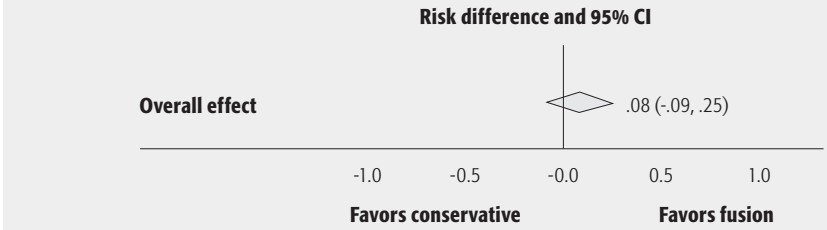
- Utility for different outcomes: Utility for different outcomes reflects the importance that an individual patient places on the outcome. This often represents a compromise among different dimensions of quality and is patient specific. For example, a patient with cervical myelopathy presented with the option of a multisegment laminoplasty and fusion will have to weigh the potential benefit of sign and symptom relief versus loss of cervical motion.

How can HTE be identified?

Examining subgroups within an RCT is a logical first step to identify HTE. Let's look at an example. Consider an RCT that compares fusion with conservative care in patients with low back pain believed to be caused by degenerative disc disease. The outcome of the study is the proportion of patients who achieve a 50% improvement in pain over baseline after 1 year. In our hypothetical example, 26% in the fusion group achieved the desired outcome compared with 18% in the conservative group with a risk difference of 8% (95% confidence interval: -9%, 25%, nonsignificant) as seen in **Fig 1**.

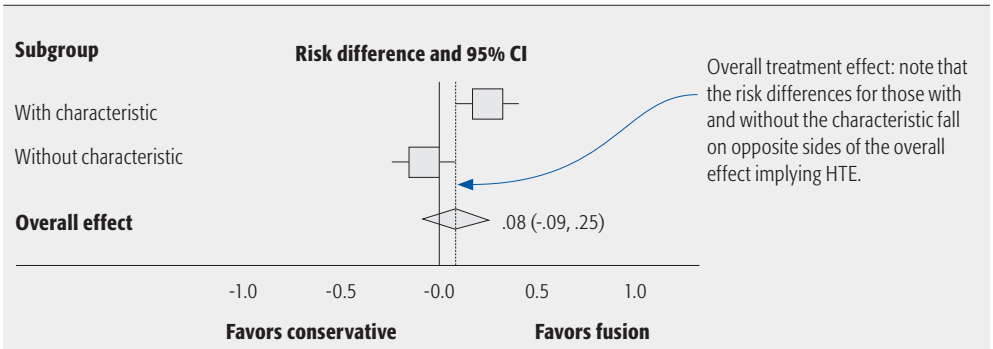
Now let's suppose there is a certain characteristic that a subgroup of patients possesses, which the investigators suspect influences the results. When the results are displayed based on the presence or absence of the characteristic, we see that those with the characteristic do better with fusion than with conservative care, while those without the characteristic do not (**Fig 2**). The differences between subgroups can be assessed statistically using a test of interaction. The interaction occurs between the treatment groups and the groups with and without the characteristic. When this happens, we say that the characteristic modifies the effect of the treatment.

Fig 1 Overall results for all patients.



The solid line represents the "null," in this case no difference in proportions of patients obtaining the desired benefit between groups. The effect of treatment for all patients is indicated by the diamond; the center, the point estimate, and its horizontal tips, the confidence interval. Statistical significance is achieved if the diamond lies completely to the left or right of the solid line. In this example, there is no statistical significance between conservative care and fusion.

Fig 2 Results for patients with and without the characteristic.



In this example, we add the risk differences stratified by the groups with and without the characteristic to the figure of overall effect. As before, there is no statistical significance between conservative care and fusion for the overall effect. The dotted line represents the point estimate of the diamond. Estimates in patients with and without the characteristic of interest on opposite sides of the dotted line suggest heterogeneity of treatment effect (HTE) and should be confirmed by a statistical test of heterogeneity.

What are the problems with identifying subgroups in RCTs?

Subgroup analyses are prone to spurious results due to the problem of multiple testing [2]. Many caution against subgroup analyses, especially post hoc comparisons [3]. Nevertheless, identification of subgroup effects in clinical trials can generate important hypotheses about potential factors that modify treatment effects. When assessing subgroups one should look for the following [3–5]:

- Statistical tests of interaction which are appropriately applied and interpreted.
- Clear description of whether the subgroup analysis was prespecified (preferred approach) or post hoc.
- An incorrect inference that a subgroup effect (interaction) is present based on separate tests of treatment effects within each level of the characteristic of interest, that is, to compare one significant and one nonsignificant *P* value [6].

What is the bottom line concerning HTE and clinical trials in spine research?

Identifying characteristics that modify treatment effects is critical to patient-centered, individualized care. Proper reporting of subgroup analysis facilitates the recognition of patients who may respond better or worse than the average. Investigators conducting

RCTs should consider characteristics that plausibly modify the effect of spine treatment, plan to conduct subgroup analyses on those characteristics and pre-state that in their protocol [3]. Considerations should be given to increasing the sample size so there is sufficient power to detect differences in subgroup analyses. While it is important to not over interpret the results of subgroup analyses, it is necessary to recognize that HTE analyses assist in hypothesis generation and aids in the design of future confirmatory studies.

Studies designed to confirm differences in treatment effects for subpopulations in turn help set the stage for this new era of "Whom to treat." Most important, identification of whom to treat may open the door for more successful patient /provider/treatment algorithms by matching up the best combinations of patient characteristics with treatment options and provider type. We could foresee a future where we more formally use consideration of specific psychosocial characteristics, phenotyping, and likely even genotyping to help us in our decision making and patient selection for a given treatment, based on evidence that is more solid. This moves us forward in our attempts to improve care and patient outcomes and helps us evolve beyond the currently "inconclusive" results from studies of chronic disease quagmires such as low back pain management [7] to clearer conclusions regarding which patients will benefit.

References

1. **Kravitz RL, Duan N, Braslow J** (2004) Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q;* 82:661–687.
2. **Brookes ST, Whitley E, Peters TJ, et al** (2001) Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess;* 5:1–56.
3. **Moher D, Hopewell S, Schulz KF, et al** (2010) CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol;* 63:e1–37.
4. **Gabler NB, Duan N, Liao D, et al** (2009) Dealing with heterogeneity of treatment effects: is the literature up to the challenge? *Trials;* 10:43.
5. **Wang R, Lagakos SW, Ware JH, et al** (2007) Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med;* 357:2189–2194.
6. **Matthews JN, Altman DG** (1996) Statistics notes. Interaction 2: Compare effect sizes not *P* values. *BMJ;* 313:808.
7. **Cahana A:** Moving forward in Concepts of Pain Relief. Spine Grand Rounds. University of Washington, Seattle, USA, February 16, 2011.