

Interpreting Negative Results From an Underpowered Clinical Trial

Warts and All

IN THE RANDOMIZED, PLACEBO-CONTROLLED TRIAL by de Haen et al¹ that appears in this issue of the ARCHIVES, the application of duct tape is compared with placebo as a treatment for common warts. The investigators recruited 103 children in the Netherlands from primary schools and assessed them for warts. Those with warts were randomized to 1 of 2 groups. One group received duct tape to apply to a designated wart, and the other group received a placebo in the form of a corn pad to apply around the wart. The primary outcome of the study was resolution of the designated wart 6 weeks after initiation of therapy. Other outcomes included change in the size of the designated wart and resolution of surrounding warts. Differences in complete resolution of the designated wart were not statistically significant. However, the investigators did find a statistically significant difference in the change in wart diameter: those treated with duct tape had a greater reduction in size compared with the placebo group. Importantly, 15% of the duct tape group stopped treatment early either because of a skin reaction to the treatment or because the tape did not stick well. Given the problems with the duct tape treatment as well as its lack of efficacy, de Haen and colleagues concluded that duct tape has a modest, nonsignificant effect as a therapy for warts.

See also page 1121

This study merits a detailed examination for several reasons. Warts are common, and a quick, effective, and inexpensive treatment is not available. Although warts are medically benign, they are unsightly and may cause a child to feel self-conscious. A recent review² concluded that the only treatment shown consistently to be effective is salicylic acid, but the treatment time is long and application is tedious. A randomized, controlled trial by Focht et al³ published in the ARCHIVES in 2002 comparing duct tape with cryotherapy showed a benefit to using duct tape. Although primary care physicians and persons with warts welcomed this news, further studies were needed to verify these results. We critiqued this study using the framework suggested by the *Users' Guide to the*

*Medical Literature: A Manual for Evidence-Based Clinical Practice*⁴ for therapeutic trials.

WERE THE STUDY PATIENTS SIMILAR TO ACTUAL PATIENTS?

A randomized, controlled trial ideally recruits participants who would most accurately reflect the population to whom the treatment would be applied in reality. For a common condition like warts, to test a treatment (duct tape) that can be purchased over the counter presents challenges to subject recruitment. Recruiting from a physician's office may not result in a representative sample, as many persons who develop warts do not seek medical attention but try other treatments at home first. Therefore, clinic-setting study participants would likely include those with warts that are more difficult to treat.

de Haen and colleagues attempted to address this limitation by recruiting from schools, which would be expected to yield a more representative, community-based sample. However, considering epidemiologic data showing that 30% of warts resolve without treatment by 32 weeks,⁵ the mean length of time that the subjects had had their warts was quite long—34.2 weeks for the experimental group and 38.5 weeks for the control group. In addition, many subjects had already tried another treatment. Therefore, the warts in the participants may be particularly resistant not only to spontaneous resolution but also to any treatment. Thus, we are doubtful that this study population is optimally representative of those who would likely consider the treatment in question. There also may have been a selection bias in recruiting, as those who were particularly motivated to enroll in the study may have been those most frustrated by warts refractory to other over-the-counter therapies. These factors may also partially explain the surprisingly low rate of spontaneous resolution within the 6-week study period as compared with the rate in other studies.²

*See the enhanced version at
<http://www.archpediatrics.com>*

**WERE PATIENTS RANDOMIZED?
WERE THE PATIENTS IN THE TREATMENT AND
CONTROL GROUPS SIMILAR WITH RESPECT
TO KNOWN PROGNOSTIC FACTORS?**

All of the 103 children were randomized to duct tape or a placebo clavi ring. Despite randomization, there were differences between the 2 groups: patients in the duct tape group reported longer wart duration on average. If we assume, in general, that older warts are more likely to spontaneously resolve, this difference in wart "age" would likely bias away from the study's null hypothesis that treatment with duct tape for 6 weeks is no different from placebo in wart resolution. However, more patients in the placebo group had already tried another treatment, which would likely bias toward null findings because such refractory warts in the placebo group would be less likely to resolve spontaneously.

**WERE PATIENTS AWARE OF GROUP
ALLOCATION? WAS THE OBSERVER AWARE
OF GROUP ALLOCATION?**

Although participants could not practically be blinded to the kind of tape they received, they were blinded to the hypothesis of the study. Randomization was blinded to the observer (wart size assessor) by having the subjects remove the tape or clavi ring prior to each assessment. The CONSORT statement⁶ recommends reporting the method used to assess blinding; however, this is often not done.⁷ To their credit, de Haen and colleagues assessed blinding in this study using 2 methods. First, the investigators asked how well the parents of participants (subject proxies) expected the treatment to work. The assumption is that those who believe they are in the experimental group will have higher expectations of treatment success. In this study, parents in both groups had similar expectations. Second, the investigators asked the observer whether she knew which kind of tape the subjects received. The observer reported that she knew the assigned group in 31% of the duct tape group compared with 17% of the placebo group, a statistically significant difference. Although the observer was incorrect about assignment more than half of the time overall, the presence of incomplete, asymmetric blinding of the observer would theoretically bias the results toward finding that duct tape was more effective. The investigators did mention that the observer was aware of the study hypothesis. Even if she were blinded to the study hypothesis, the observer could still have subconsciously formed hypotheses of her own, which could potentially bias the results.

**WAS FOLLOW-UP COMPLETE?
WERE PATIENTS ANALYZED IN THE GROUPS
TO WHICH THEY WERE RANDOMIZED?**

In contrast to the prior randomized, controlled trial of duct tape for warts by Focht et al,³ follow-up by de Haen and colleagues was complete, which improves the validity of these results. All of the patients were accounted for at the end of the study, and follow-up was conducted in person rather than by telephone for all of the subjects. Furthermore, all of the participants were analyzed in the

groups to which they were randomized, and this even included those who did not complete treatment. This intention-to-treat analysis included everyone regardless of adherence to treatment and was designed to produce conclusions that are more applicable to real-life situations.

An intention-to-treat analysis reflects the potential real-world experience in which not all of the patients adhere to treatment recommendations. However, including in the analysis those who did not finish the course of treatment could dilute a positive effect of the treatment. For example, if all of the 8 subjects in the duct tape group whose warts resolved had completed treatment, analyzing treatment completers as a subgroup may have yielded a different result and led to different conclusions. If duct tape for warts was effective for this subgroup, improving the stickiness of the tape and treating concomitant eczema more aggressively may be more appropriate than entirely dismissing duct tape as a potentially beneficial treatment. Although such post hoc analyses are methodologically less appealing than approaches chosen by investigators prior to a study, they can potentially inform future studies as well as clinical management.

STATISTICAL POWER AND TYPE II ERROR

de Haen and colleagues found no statistical difference in wart resolution between duct tape and placebo. As these were negative results, it is essential to consider the likelihood of a type II error, or that the investigators erroneously concluded that duct tape was ineffective. Did the study have enough power to detect a clinically meaningful difference between the 2 groups, and how did the investigators decide how big that difference should be?

The primary outcome in this study was binary, meaning that there were 2 possible and distinct outcomes: either the wart resolved or it did not. To calculate how many subjects are needed to detect a difference in a binary outcome, one must estimate how many subjects in each group will have the desired outcome. Previous studies⁸ have suggested that 30% of warts resolve after 10 weeks with no treatment, and de Haen and colleagues decided that an acceptable treatment alternative should be 30 percentage points better than the natural history of warts.¹ Thus, they calculated how many subjects were needed to detect a significant difference if 30% of warts in the placebo group resolved and 60% of warts in the duct tape group resolved.

However, de Haen and colleagues misjudged 2 critical pieces of information in calculating their desired sample size. First, for a 6-week—rather than 10-week—follow-up, they should have anticipated a lower rate of wart resolution because of the shorter time frame. Second, they should have elaborated more on how they concluded that a 30-percentage-point difference was clinically meaningful, especially because many persons with warts may consider a smaller effect size beneficial enough—and the treatment accessible enough—to try the duct tape treatment on their own.

If de Haen and colleagues had recruited a larger sample to detect a smaller but still clinically relevant effect, they may have reached different conclusions with the very same

Table. Precision of Estimates of the Effect of Duct Tape vs Placebo for Wart Therapy

Treatment	Patients, No.	Treated Warts That Resolved, %	95% CI*
Duct tape	51	16	7.0-28.6
Placebo	52	6	1.2-15.9

Abbreviation: CI, confidence interval.

*The CIs were calculated from data provided by de Haen et al.¹

effect size they described. With 6% spontaneous resolution in the placebo group and a difference of only 10 percentage points between duct tape and placebo, de Haen and colleagues had insufficient power (only 26%) to reject their null hypothesis of no difference between duct tape and placebo. In contrast, if they assumed a priori 30% spontaneous resolution in placebo and a 20-percentage-point meaningful difference, they would have needed to recruit 184 patients evenly split between the groups to achieve 90% power to reject the null hypothesis. If they had such a sample and had identical findings to what they described (ie, lower spontaneous resolution rate and smaller treatment effect than anticipated),¹ they would have had 53% power to reject the null hypothesis. If instead they had assumed a priori 30% resolution and a 15-percentage-point difference, their target sample size would have been about 350 children. A sample of this size would have given them more than 80% statistical power; importantly, the same results they described would have led them to *reject* the null hypothesis at $P < .05$. Although such a study would have been more expensive, it would have provided a better test of the null hypothesis.

HOW PRECISE WAS THE ESTIMATE OF THE TREATMENT EFFECT?

In a randomized, controlled trial, the measured effect on the subjects is an estimate of what the true effect would be if the treatment were applied to a larger population. The precision of the estimate is important in interpreting these results. Precision is estimated by confidence intervals. The 95% confidence intervals give us parameters with which we can say that if we repeat the same study 100 times, the result would fall within these parameters 95% of the time. Although de Haen and colleagues did not report confidence intervals, we calculated them to provide insights into the reported data (**Table**).

These confidence intervals are somewhat wide, largely attributable to the small sample noted earlier. There is considerable overlap with these confidence intervals, consistent with the failure by de Haen and colleagues to reject the null hypothesis of no difference between 6 weeks of treatment vs placebo.

WERE ALL CLINICALLY IMPORTANT OUTCOMES CONSIDERED?

As a secondary outcome, de Haen and colleagues measured the change in the diameter of the wart over the

6-week study period and found significant differences between the groups. This is an important finding, especially given the short observation period. The warts treated with duct tape were improving more rapidly. It is reasonable to speculate that if the study continued, the warts treated with duct tape would be likely to resolve more quickly than the placebo-treated warts, and at some point, meaningful differences in resolution rates that de Haen and colleagues anticipated would be achieved.

CONCLUSIONS

The negative study published by de Haen and colleagues is illustrative for several reasons. First, it allows us to examine the potential effect of selection bias in recruiting subjects for a study of a common condition. Second, de Haen and colleagues assessed how well the subjects and observers were blinded, which provides the opportunity to discuss this very important but often underreported component of randomized, controlled trials. Third, this study shows possible limitations of intention-to-treat analyses and the potential benefit of as-treated subgroup analyses in randomized, controlled trials to assess the treatment effects when the sample is heterogeneous with respect to treatment exposure. Fourth and perhaps most importantly, the study illustrates the perils of study design with respect to sample size. Although investigators must always balance scientific goals with fiscal and logistic constraints, the assumptions by de Haen and colleagues regarding the spontaneous wart resolution rate and a meaningful clinical difference were so substantively different from what they actually found that they were left underpowered to assess their study hypotheses.

In summary, this study tests an inexpensive treatment for a common condition, and the results contradict an earlier randomized, controlled trial that had flaws of its own.³ However, several methodological limitations in the study by de Haen and colleagues lead us to question the investigators' conclusions that the effects of duct tape were not significant. Further studies that address the limitations of these extant studies are needed before such definitive conclusions can be drawn.

Jeanne Van Cleave, MD
Alex R. Kemper, MD, MPH, MS
Matthew M. Davis, MD, MAPP

Correspondence: Dr Davis, Division of General Pediatrics, University of Michigan, 300 N Ingalls Bldg, Room 6C23, Ann Arbor, MI 48109-0456 (mattdav@med.umich.edu).

Author Contributions: *Study concept and design:* Davis and Van Cleave. *Analysis and interpretation of data:* Davis and Kemper. *Drafting of the manuscript:* Davis and Van Cleave. *Critical revision of the manuscript for important intellectual content:* Davis and Kemper. *Statistical analysis:* Davis and Van Cleave. *Administrative, technical, and material support:* Davis. *Study supervision:* Davis.

Financial Disclosure: None reported.

REFERENCES

1. de Haen M, Spigt MG, van Uden CJT, van Neer P, Feron FJM, Knottnerus A. Efficacy of duct tape vs placebo in the treatment of verruca vulgaris (warts) in primary school children. *Arch Pediatr Adolesc Med.* 2006;160:1121-1125.
2. Gibbs S, Harvey I, Sterling JC, Stark R. Local treatments for cutaneous warts [update of *Cochrane Database Syst Rev.* 2001;(2):CD001781]. *Cochrane Database Syst Rev.* 2003;(3):CD001781. doi:10.1002/14651858.CD001781. Processed June 29, 2006.
3. Focht DR, Spicer C, Fairchok MP. The efficacy of duct tape vs cryotherapy in the treatment of verruca vulgaris (the common wart). *Arch Pediatr Adolesc Med.* 2002;156:971-974.
4. Guyatt G, Rennie D. *Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* Chicago, Ill: American Medical Association; 2002.
5. Massing AM, Epstein WL. Natural history of warts: a two-year study. *Arch Dermatol.* 1963;87:306-310.
6. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134:663-694.
7. Boutron I, Estellat C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable. *J Clin Epidemiol.* 2005;58:1220-1226.
8. Gibbs S, Harvey I, Sterling J, Stark R, Altman DG. Local treatments for cutaneous warts: systematic review. *BMJ.* 2002;325:461.

Announcement

Topic Collections. The *Archives* offers collections of articles in specific topic areas to make it easier for physicians to find the most recent publications in a field. These are available by subspecialty, study type, disease, or problem. In addition, you can sign up to receive a Collection E-Mail Alert when new articles on specific topics are published. Go to <http://archpedi.ama-assn.org/collections> to see these collections of articles.