In the past two decades, two landmark randomized controlled trials (RCT) have been completed among individuals with sickle cell disease (SCD), the Multi-center Study of Hydroxyurea (MSH) trial and the Stroke Prevention (STOP) trial. The MSH trial tested the hypothesis that hydroxyurea will reduce the frequency of painful episodes for adults with hemoglobin SS who had a history of 3 or more painful episodes per year. The STOP trial tested the hypothesis that among children with hemoglobin SS and an elevated transcranial Doppler (TCD) velocity measurement, blood transfusion therapy would decrease the risk of an initial stroke. After completion, both trials have defined standard care for individuals with hemoglobin SS. The purpose of this review is to examine the limitations of the MSH and STOP trials. In the context of these trials, we will examine the effects of narrow inclusion criteria that primarily include participants with hemoglobin SS and secondary analyses that are prone to false-positive results. In addition, we describe how after publication of these two trials use of hydroxyurea and TCD assessment has drifted towards a standard practice without evidence of therapeutic efficacy among groups that were excluded from the trials. Finally, we suggest that rigorously conducted RCTs or at the minimum multicenter observation studies with strong methodology should be performed in these excluded subgroups to confirm a benefit of hydroxyurea or TCD measurement.

Clinical Trials
In the hierarchy of study design, randomized controlled trials (RCTs) are considered the best method to rigorously test a hypothesis and subsequently change clinical practice. As opposed to observational studies, the experimental design of RCTs most effectively minimizes the influence of confounding variables to ascertain a reliable treatment effect. However, even well-designed and executed RCTs have limitations. For this review, we will discuss three limitations of clinical trials (narrow spectrum of inclusion criteria, over-reliance on secondary analyses, and extrapolating trial results to other clinical settings) in the context of the Multi-center Study of Hydroxyurea (MSH) trial and the Stroke Prevention (STOP) trial (Table 1). Our objectives are to provide insight into the limitations of RCTs so that sickle cell disease (SCD) practitioners may better apply the findings of RCTs to their patient population and to identify areas of investigation requiring additional research.

The first limitation of RCTs pertinent to practitioners caring for individuals with SCD is narrow inclusion criteria. Limiting a clinical trial to a homogeneous group of participants lessens potential confounders, but the end result may be a study group that is representative of only a minority of the targeted patient population.

Second, subgroup analyses or secondary analyses of an RCT are often considered to have the same weight and
influence as the primary analysis. Clinical trials are designed to address a primary hypothesis with a primary analysis. Thereafter, a therapeutic benefit can be sought in any number of secondary analyses. Results from these analyses are inherently less reliable due to a risk of type I (false-positive) errors. By convention, a significance level set at 5% indicates the chance of falsely concluding a statistically significant relationship exists. If multiple comparisons are made, the risk of incorrectly rejecting the null hypothesis increases precipitously. The relationship between the probability of a type I error and the number of comparisons is as follows—probability of type I error = 1 – (1-α)^n (α = significance level, most commonly .05, n = number of comparisons). Thus, if 3 separate secondary analyses are performed with an α of .05, the type I error rate increases to 14%. Given the complex nature of clinical trials designed with the intent of addressing a single primary hypothesis but collecting vast amounts of data, ample opportunities exist for secondary analyses and corresponding type I errors.

One strategy to lessen the probability of type I errors in secondary analyses is to use a more stringent alpha level (e.g., <.01). While a reduction in the alpha level may not be necessary for a priori secondary hypotheses, adjustments should strongly be considered for post-hoc analyses that involve multiple comparisons. However, these strategies are rarely used. Most commonly, secondary analyses are reported as a primary hypothesis in the same or separate manuscript without adjustment of the level of significance to decrease the probability of a type I error. Without adjusting the level of significance from .05 to a lower threshold, secondary analyses lead to false-positive results. At the very least, secondary analyses should be identified as such and a rationale discussion made as to the pros and cons for why no adjustment of the level of significance was done.

A third limitation of RCTs is not the results of the trial per se, but new practice patterns emerging after the completion of the trial. Often, clinicians extrapolate the data from an RCT to other clinical situations that were not originally included in the trial. For example, despite the potential differences in dosing, safety, and efficacy between drug trials in adults when compared with children, pediatricians are often forced to rely on data from studies performed in adults to make treatment decisions in children. Based on these known limitations, the US Food and Drug Administration (FDA) developed a program that offered incentives to pharmaceutical companies to perform studies in children. From 1998 to 2004, 253 of these studies were conducted, yet only 50% of the pediatric trials resulted in a new indication for the pediatric setting, highlighting the limitation of therapeutic trials in adults being the basis for treatment in children. Rather than accepting adult SCD trials as sufficient evidence for pediatric practice or relying on pediatric trials for management of adults with SCD, pediatric- and adult-specific trials in SCD should dictate practice patterns of treating hematologists.

Limiting Inclusion to Individuals with Sickle Cell Anemia or \( S^\beta^0 \) Thalassemia in Clinical Trials

Most RCTs in SCD have limited the inclusion criteria by hemoglobin phenotype. Both the MSH and STOP trials limited the participants to individuals with hemoglobin SS (and \( S^\beta^0 \)-thalassemia in the STOP trial) and excluded individuals with hemoglobin SC. Several single-institution, single-arm studies have suggested a therapeutic efficacy of treatment with hydroxyurea among individuals with hemoglobin SC; however, no RCT has been completed to assess the safety and efficacy of hydroxyurea among individuals with hemoglobin SC.
individuals with this SCD phenotype. Given the marked difference in incidence of pain for persons with SC disease when compared with SS disease—0.4 episodes per year and 0.8 episodes per year, respectively—the potential efficacy and risk benefit ratio of hydroxyurea requires a formal evaluation among individuals with SCD.\textsuperscript{12} Even with limited evidence, treatment with hydroxyurea for individuals with hemoglobin SC has received tacit support within the sickle cell community and is slowly emerging as standard clinical practice for individuals with hemoglobin SC and multiple painful episodes. Currently, a National Institute of Health (NIH)–National Heart, Lung and Blood Institute (NHLBI) phase 2 clinical trial is being conducted among individuals with hemoglobin SC disease that includes hydroxyurea and magnesium pidolate (W. Wang, personal communication).

Similar to the MSH trial, participants with hemoglobin SC were excluded from the STOP trial. The prevalence of overt strokes among children with hemoglobin SC is substantially lower than in children with SS, 0.01 versus approximately 0.04.\textsuperscript{11,13} but still higher when compared with children without SCD.\textsuperscript{14} No RCT has been performed, and only scant data exist describing the relationship between transcranial Doppler (TCD) measurement and the risk of strokes among children with hemoglobin SC. Despite this limitation, at least one report has used TCD measurement to assess velocity in the middle cerebral artery in this population.\textsuperscript{15} Although this report does not advocate screening individuals with hemoglobin SC disease, no explicit guidelines state that TCD measurement should be limited to patients with hemoglobin SS or S\textsuperscript{β} thalassemia. Recommendations from the American Academy of Neurology advocate TCD screening in all children with SCD regardless of phenotype.\textsuperscript{16} However, in the event that the TCD measurement is greater than 200 cm/second in a child with SCD, no data exist to estimate the incidence of an initial stroke without blood transfusion or the benefit of blood transfusion therapy in preventing primary strokes. The relative merits of screening for elevated TCD measurements among children with hemoglobin SC is worthy of formal evaluations, else this practice will continue without evidence.

Over-Reliance on Secondary Analyses
Both the MSH and the STOP trials collected significant data that allowed for secondary analyses, analyses that were not based on the primary hypothesis. Due to the strength of the original trials, these secondary studies are often given the same weight of credibility as the primary analysis. However, when closely evaluated, these secondary analyses are rarely performed with the same rigor as the primary analysis. Specifically, intention-to-treat analyses are often not included, and the adjudication process for the secondary outcomes are commonly not as well defined. Further, these secondary analyses are often unplanned and adjustment in the threshold for significance, from .05 to a lower level, is commonly not done (Table 2).

Following completion of the MSH trial, participants were given the option to continue, discontinue, or initiate hydroxyurea therapy if they were in the observation arm. Observational data was then prospectively collected for up to 9 years. A total of 299 participants elected to receive hydroxyurea, and 233 individuals were able to be evaluated. Using an intention-to-treat analysis, the original placebo group was compared to the hydroxyurea group, and no difference in survival (\(P = .35\)) was noted between the two assigned groups. However, a 40\% reduction in mortality was found when individuals were compared based on total duration of hydroxyurea therapy without attention to the original study grouping (\(P = .04\)).

Several limitations weaken the evidence that hydroxyurea increases overall survival in adults. First, the follow-up was an observational study, not a randomized trial. Study participants self-selected whether they wanted to continue, discontinue, or start hydroxyurea treatment. The process of self-selection may have resulted in nonsystematic bias. For example, individuals who elected to continue receiving hydroxyurea or start therapy could be less likely to have a comorbid condition such as renal failure, a risk factor for premature death\textsuperscript{17} and a relative contra-indication for the use of hydroxyurea; or they may be more likely to have private insurance or come from a higher social economic group. Any one or all of these factors may contribute to an improvement in overall survival in those who elected to take hydroxyurea when compared with those who did not. Second, no direct evidence was available that individuals actually took the medication. Specifically, the measures that were used in the MSH study to assess adherence to hydroxyurea were not used in the observational study, and no other evidence was offered to support that participants were compliant with therapy throughout the observational study period.

Perhaps the strongest limitation of the post-hoc analysis of the MSH survival data was the exploratory nature of the analysis coupled with the marginal level of significance for the survival analysis (\(P = .04\)). The authors of the primary analysis were aware of the problem associated with multiple comparisons and planned for an adjustment of the level of significance for the secondary analysis. In the primary manuscript the following statement was included in the Statistical Analysis section, “To adjust for multiple tests of the data in secondary analyses, two-sided tests with \(P\) values between .01 and .001 were considered to provide some evidence of significance differences between the groups, and the tests with \(P\) values below .001 were considered to provide strong evidence of such differences.”\textsuperscript{18} Despite recognition of the risk of a type I error, the post-hoc analysis did not adhere to previously stated level of significance. If a significance level of .001 or even .01 was adhered to as designated in the primary manuscript, the results of the trial would have received far less attention. Further, in recognition of the marginal statistical signifi-
cance of the findings, the data may have been used as pre-
liminary evidence for a formal RCT to test the hypothesis
that hydroxyurea improves overall survival. However, 4
years after the publication of the post-hoc analysis, con-
sensus now exists that hydroxyurea therapy prolongs sur-
vival in adults with hemoglobin SS. A future trial to for-
mally test if hydroxyurea prolongs survival in adults, and
if so, whether the risks outweigh the benefits, is unlikely to
be initiated due to the lack of equipoise in the sickle cell
community. There are published guidelines that can help cli-
nicians determine the reliability of secondary analyses5 (Table
3). In general, secondary analyses are most useful for generat-
ing new hypotheses, or replicating results.

Another example of a secondary analysis of a land-
mark study occurred in the STOP trial, when findings were
published demonstrating that children who received trans-
fusion therapy had a significant decrease in the number of
pain and acute chest syndrome (ACS) episodes. Miller et al
used an intention-to-treat analysis to show a reduction in
the frequency of ACS episodes ($P = .0027$); however, this
analysis did not demonstrate significant reductions for pain
($P = .13$). Thereafter, the authors limited the analysis to

Table 2. Secondary analyses of the Multi-center Study of Hydroxyurea (MSH) trial and the Stroke Prevention (STOP) trial.

<table>
<thead>
<tr>
<th>Secondary analysis</th>
<th>Title</th>
<th>Level of significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSH</td>
<td>Hydroxyurea and sickle cell anemia: effect on quality of life</td>
<td>.01</td>
</tr>
<tr>
<td>Ballas et al46 (2006)</td>
<td>Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: risks and benefits up to 9 years of treatment</td>
<td>.05</td>
</tr>
<tr>
<td>Steinberg et al37 (2003)</td>
<td>Cost-effectiveness of hydroxyurea in sickle cell anemia</td>
<td>.05</td>
</tr>
<tr>
<td>Moore et al38 (2000)</td>
<td>Fetal hemoglobin in sickle cell anemia: determinants of response to hydroxyurea</td>
<td>.01</td>
</tr>
<tr>
<td>Steinberg et al39 (1997)</td>
<td>Enhanced blood flow velocity in the anterior cerebral artery and stroke risk in sickle cell disease: extended analysis from the STOP trial</td>
<td>.05</td>
</tr>
<tr>
<td>Lee et al41 (2006)</td>
<td>Regular transfusion lowers plasma free hemoglobin in children with sickle cell disease at risk for stroke</td>
<td>.05</td>
</tr>
<tr>
<td>Lezcano et al42 (2006)</td>
<td>Effect of long-term transfusion on growth in children with sickle cell anemia: results of the STOP trial</td>
<td>.05</td>
</tr>
<tr>
<td>Wang et al43 (2005)</td>
<td>Magnetic resonance angiography in children with sickle cell disease and abnormal transcranial doppler ultrasonography findings enrolled in the STOP study</td>
<td>.05</td>
</tr>
<tr>
<td>Abboud et al44 (2004)</td>
<td>Stroke and conversion to high risk in children screened with transcranial doppler ultrasound during the STOP trial</td>
<td>Not stated</td>
</tr>
<tr>
<td>Jones et al46 (2004)</td>
<td>Alpha thalassemia is associated with decreased risk of abnormal transcranial doppler ultrasonography in children with sickle cell anemia</td>
<td>.05</td>
</tr>
<tr>
<td>Hsu et al47 (2003)</td>
<td>Longitudinal changes in ferritin during chronic transfusion: a report from the STOP trial</td>
<td>Not stated</td>
</tr>
<tr>
<td>Files et al48 (2002)</td>
<td>Impact of chronic transfusion on incidence of pain and acute chest syndrome during the STOP trial in sickle cell anemia</td>
<td>Not stated</td>
</tr>
<tr>
<td>Miller et al49 (2001)</td>
<td>Multicenter comparison of magnetic resonance imaging and transcranial doppler ultrasonography in the evaluation of the central nervous system in children with sickle cell disease</td>
<td>Not stated</td>
</tr>
</tbody>
</table>

Table 3. An example of guidelines for interpreting secondary or subgroup analyses using the mortality analysis from the Multi-Center Study of Hydroxyurea (MSH) trial (adapted from Oxman, et al).5

1. Is the magnitude of the difference clinically important? Yes, 40% reduction in mortality.
2. Was the difference statistically significant? No, $P$ values were not less than .01.
3. Did the hypothesis precede rather than follow the analysis? No, mortality was not mentioned in the primary manuscript.
4. Was the subgroup analysis one of a small number of hypotheses tested? No, several were tested (for example, episodes of acute chest syndrome, time to pain episodes, transfusion requirements).
5. Was the difference suggested by comparisons within rather than between studies? Yes, comparisons were within the study.
6. Was the difference consistent across studies? There are no other studies to compare.
7. Is there indirect evidence that supports the hypothesized difference? Yes, it makes biologic sense that hydroxyurea could improve survival.
children in the treatment arm who had been adherent to transfusion therapy. When the compliant group was analyzed, the authors were able to demonstrate a decrease in the number of pain episodes versus the nontransfused group ($P = .014$). Despite the limitations of the secondary analysis, the results are consistent with a previous study and confirm the prevailing belief that blood transfusion prevents subsequent pain and ACS episodes among individuals with hemoglobin SS disease. Still, the benefit of blood transfusion therapy to prevent pain and ACS episodes when measured against the risks has not been formally defined.

**Extrapolating the Results of the Trial to Other Clinical Settings Resulting in a Clinical Drift to Change Practice Without Direct Evidence**

In the MSH trial, there was significant evidence for a reduced rate of pain and ACS among adults. Subsequently, these results have been extrapolated as evidence to treat children with multiple painful episodes or repetitive ACS episodes. The most rigorous hydroxyurea study performed in pediatrics was a phase 2 study funded by NHLBI that enrolled 84 children between the ages of 5 and 15 years. This study and others have demonstrated that when hydroxyurea was administered to children, similar hematologic responses to those seen in adults occur, such as a rise in fetal hemoglobin and mean cell volume, along with a reduction in WBC and reticulocyte count. Further, no significant short-term toxicity was demonstrated. In a prospective infant cohort study ($n = 21$) with a mean follow-up of 4.9 years, Hankins et al reported a reduction in morbidity for children treated with hydroxyurea. Episodes of ACS occurred 7.5 times per 100 patient-years in the group treated with hydroxyurea compared with 24.5 episodes per 100 patient-years observed in historic controls. Further, no significant short-term toxicity was demonstrated. In a prospective infant cohort study ($n = 21$) with a mean follow-up of 4.9 years, Hankins et al reported a reduction in morbidity for children treated with hydroxyurea.

Another example of clinical practice that has drifted from the original trial is the timing of TCD assessment. The STOP trial had well-defined criteria as to the clinical setting in which measurements could be obtained. These settings included, but were not limited to, children without evidence of an acute illness. Despite the established criteria, clinicians have decided to perform TCD measurements when patients are hospitalized and acutely ill. The relationship between pain or ACS episodes and TCD velocity has not been studied in a clinical trial setting. However, the increase in TCD velocity associated with a lower hemoglobin has been well documented. The expected drop in hemoglobin among children with pain or ACS episodes would likely result in a higher-than-expected baseline measurement. Thus, the practice of obtaining TCD measurements while the patient is acutely ill may lead to a subgroup with falsely elevated velocities.

**Rationale for Additional and Specific Clinical Trials for Hemoglobin SC Disease and Children with SCD**

This review highlights the need for RCTs as a cornerstone to improve the care of individuals with SCD. The findings from RCTs are the highest form of evidence for a treatment intervention, and these findings rightfully direct clinical practice. Limited resources do not permit RCTs to address every major clinical problem in SCD. In the absence of a RCT, rigorously designed observational studies (cohort or case-control design) can provide reasonable evidence to guide clinical practice. Nevertheless, additional RCTs are needed for more precise recommendations for all individuals with SCD regardless of phenotype or age group.

A major, but appropriate, limitation in the two landmark SCD clinical trials is the exclusion of individuals with hemoglobin SC. Consequently, results from these trials do not provide evidence that individuals with hemoglobin SC disease should be treated in the same manner as
their hemoglobin SS counterparts. Based on the estimate that 70,000 Americans have SCD, at least 20,000 have hemoglobin SC disease, approximately 10 times more individuals than with β-thalassemia major, yet no formal phase 3 trial has been completed among individuals with hemoglobin SC disease. Rather than making inferences from trials that are designed for participants with hemoglobin SS, resources should be allocated to answer major clinical questions for both individuals with hemoglobin SS and those with hemoglobin SC.

As well, drugs such as hydroxyurea with proven efficacy in adults with SCD should be investigated thoroughly in children. In the general population, the vast majority of drugs for children lack appropriate dosing, safety, and efficacy data. Given strong evidence indicating different toxicities and benefit for drugs in children compared with adults, pediatric therapeutic trials in SCD are necessary after completion of adult studies to appropriately assess investigational agents. When there is limited market value in conducting such trials, the NIH becomes one of the few alternatives available to sponsor such RCTs. The lack of sufficient efficacy data for the use of hydroxyurea in children, coupled with the limited marketing value of hydroxyurea, underscores the importance of conducting trials that ultimately must be sponsored by the NIH, or else they are unlikely to ever be completed.

For the current and future generation of individuals with SCD, RCTs that build on existing knowledge and expand to more diverse subgroups of individuals with SCD are required to enhance our knowledge about the best strategies to provide care for this vulnerable population.

Acknowledgments

Supported in part by the National Institutes of Health, National Heart, Lung and Blood Institute, RO1 HL079937 (MRD), K12HL08710 (JJF).

Correspondence

Michael R. DeBaun, MD MPH, Washington University School of Medicine in St Louis, CB 8519, St Louis, MO 63108; phone (314) 286-1174; fax (314) 286-2609; debaun_m@kids.wustl.edu

References

27. Ross CS, Hennessy JM, Nelson SC. Subarachnoid


