Hyperthermic Isolated Limb Perfusion for Extremity Soft Tissue Sarcomas: Systematic Review of Clinical Efficacy and Quality Assessment of Reported Trials

N.H. TRABULSI, MD, L. PATAKALVI, M.O. NASSIF, MD, R.E. TURCOTTE, MD, FRCS, A. NICHOLS, PhD, and A.N. MEGUERDITCHIAN, MD, FRCS, FACS*

1Department of Clinical Epidemiology, McGill University, Montreal, Canada
2Department of Surgery, King Abdulaziz University Hospital, Jeddah, Saudi Arabia
3Department of Surgery, McGill University, Montreal, Canada
4Department of Orthopedic Surgery, McGill University, Montreal, Canada
5Urban Institute, Washington, District of Columbia

Background and Objectives: Extremity soft tissue sarcomas (STS) are managed with radiotherapy and limb-sparing surgery however aggressive or recurrent cases require amputation. Hyperthermic isolated limb perfusion (HILP) has been proposed as an alternative. Our aim was to systematically review phase II HILP trials, assess tumor response, limb salvage (LS), and quality of scientific publications on this technique.

Methods: We conducted a literature search of electronic databases (MEDLINE, EMBASE, Scopus, Cochrane Library) and clinical trial registries for phase II HILP trials on non-resectable extremity STS. Outcomes of interest were complete response (CR), partial response (PR), and LS rates. Quality of published trials was assessed using a quality checklist.

Results: Of 518 patients across 12 studies, 408 had some response (CR or PR), and 428 had the limb spared. Median CR, PR, and LS rates were 31%, 53.5%, and 82.5%, respectively. Median Wieberdink loco-regional toxicity rates were 3.8%, 45.5%, 17%, 1%, and 0% for levels 1–5, respectively. No trial fulfilled either all ideal or essential quality criteria. Seven trials did not include statistical methodology.

Conclusion: HILP seems effective in treating advanced extremity STS. However, poor publication quality hinders results validity. Technical and methodological standardization, well-designed, multi-institutional trials are warranted.


KEY WORDS: regional chemotherapy; limb; advanced disease

INTRODUCTION

Soft tissue sarcomas (STS) are a rare group of tumors, representing 1% of adult malignancies. The extremities are the most commonly involved sites, accounting for 50–60% of cases [1,2]. The treatment of extremity sarcomas has shifted to more conservative approaches in recent years, with the understanding that amputation does not enhance survival in patients with large (>5 cm) high-grade sarcomas [3]. Limb-sparing resections can be achieved in as many as 90% of extremity STS with multimodal therapy (surgery and radiation therapy) [4]. However, for locally advanced or recurrent STS, it is often impossible to completely resect the tumor without performing an amputation or leaving the patient with a functionless limb [5,6].

Hyperthermic isolated limb perfusion (HILP), first described by Creech et al. [7], has been described as an alternative for patients with extremity STS. This technique is based on a few basic principles [8–10]. Circulation to the extremity is isolated from the rest of the body by placing catheters into the extremity’s main artery and vein. High-dose chemotherapy is administered to the extremity through a pump, and leakage to systemic circulation is limited by application of a tourniquet proximal to the administration site. The temperature within the isolated circuit is increased to make the tumor cells more vulnerable to the chemotherapy. The rationale behind this technique is that systemic toxicity will be avoided and exposure to high-dose cytotoxicity will be enhanced in a limited field [11]. A variety of chemotherapeutic agents have been used, including melphalan, doxorubicin, tumor necrosis factor (TNF)-alpha, and interferon gamma [8–10]. Many studies have assessed the effectiveness of HILP for tumor response, limb-sparing rates, complications, systemic toxicity, mortality, and morbidity. However, no phase III trial has been published to date.

The purpose of this study was to systematically review phase II trials related to HILP as an effective treatment modality of locally advanced STS of the extremities, with concurrent assessment of the quality of these scientific publications. We also aimed to summarize the findings of these phase II trials in order to assess whether or not there is enough promise to consider phase III trials.

METHODS

Search Objectives

The main objective of this review was to conduct a systematic and comprehensive literature search to identify published trials that

Abbreviations: STS, soft tissue sarcoma; HILP, hyperthermic isolated limb perfusion; TNF, tumor necrosis factor; WHO, World Health Organization; CR, complete response; PR, partial response; LS, limb salvage; RCTs, randomized controlled trials.

Grant sponsor: Cedars Cancer Institute (Montreal, Canada).
*Correspondence to: A.N. Meguerditchian, MD, FRCS, FACS, Surgical Oncology, McGill University Health Centre, 1140 Pine Avenue West, Montreal, Canada. Fax: 514-843-1633. E-mail: ari.meguerditchian@mcgill.ca

Received 5 January 2012; Accepted 29 May 2012
DOI 10.1002/jso.23200
Published online in Wiley Online Library (wileyonlinelibrary.com).

© 2012 Wiley Periodicals, Inc.
2 Trabulsi et al.

had assessed HILP as an effective treatment modality for locally advanced STS of the extremities.

**Protocol and Registration**

The protocol is available for review at [http://www.crd.york.ac.uk/PROSPERO](http://www.crd.york.ac.uk/PROSPERO). The registration code is CRD42011001390. A health science librarian was involved in structuring the protocol and search strategy.

**Search Strategy**

From June 20 to 24, 2011, we conducted a comprehensive literature search that included electronic databases (MEDLINE, EMBASE, Scopus, and Cochrane Library), an electronic conference website (NLM gateway meeting abstracts), and clinical trial registries (clinicaltrials.gov and International Clinical Trials Registry Platform; World Health Organization [WHO]). Literature published to June 2011 and relevant reference lists of studies identified in the electronic search were retrieved. Journals of interest (Sarcoma Journal and the International Journal of Hyperthermia) were indexed at EMBASE. The search terms for the MEDLINE database were as follows: (“sarcoma” AND (“extremity” OR “limb” [TIAB]) AND “adult”) AND (“melphalan” OR “doxorubicin” OR “dactinomycin” OR “tumor necrosis factor” OR “drug therapy” OR “chemotherapy, cancer, regional, perfusion” OR “hyperthermia, induced” OR “perfusion” OR “infusion” [TIAB] OR “isolated” [TIAB]). We limited our searches to adults and humans.

**Study Selection**

After completion of all searches, duplicates were removed. Two reviewers independently screened studies initially based on the title, key words, and abstract of the retrieved record to exclude non-relevant, non-English, non-human, and non-adult studies. Articles for full text review were assessed independently by two authors using a checklist for study selection (Appendix 1). Studies that did not meet the inclusion criteria were discarded during the initial review. When uncertainty existed, we retrieved and assessed the full-text studies. A third author resolved differing opinions.

**Inclusion criteria.** Studies were included if they satisfied all of the following criteria.

**Study design.** Publications to be included were trials in which the main purpose was to assess the efficacy of isolated limb perfusion (phase II trials, or phase I trials that had looked at the maximum tolerated dose and efficacy as a primary or secondary outcome). We reviewed only those trials in which the intervention was under controlled settings, the patients included received standardized therapeutic regimes, and assessment methods and follow-up were standardized. Whenever the same trial was reported (published) multiple times, we included only the most recent report in order to avoid duplication of patients in overlapping series.

**Study subjects.** Studies involved human subjects, adults 18 years and over, with unresectable, histologically confirmed STS of the extremities.

**Interventions.** Interventions included HILP for advanced STS of the extremities. Trials had to report complications and technique in detail (achievement of hyperthermia, isolation of the limb, and creation of a circuit).

**Outcomes.** Complete response (CR), partial response (PR), and limb salvage (LS) rates had to be reported in the trial. According to the WHO handbook of objective tumor response assessment of measurable disease [12], these clinical outcomes are measured as follows:

- **Complete response:** The disappearance of all known disease, determined by two observations not less than 4 weeks apart.
- **Partial response:** 50% or more decrease in total tumor size of the lesions, which have been measured to determine the effect of therapy by two observations not less than 4 weeks apart. In addition, there can be no appearance of any new lesions or progression of any lesion.
- **Limb salvage:** We did not find standardized criteria to determine LS occurrence (in terms of time frame), nor had the trials that we identified determined any such criteria. All trials considered that LS was achieved if the disease was addressed without the need of amputation or limb-mutilating surgery.

**Exclusion criteria.** Prospective and retrospective cohort studies, case reports, and series and commentaries about HILP were to be excluded. Trials whose main purpose was to assess HILP as a palliative procedure were also to be excluded.

**Data Collection Process and Data Items**

Two reviewers extracted data that met the inclusion criteria by using a prespecified extraction form containing the following information: settings, number of patients included/analyzed, age median (or mean), median follow-up, technique used, chemotherapeutic agents, TNF doses (high vs. low), and whether patients within a trial received different regimes.

We also collected reported outcomes: CR, PR, and LS rate. Whenever the CR and PR were not reported per the WHO or the National Cancer Institute criteria, the reported results were standardized to meet those criteria. Since this systematic review aimed at examining the efficacy of a relatively new intervention, it was critical to look at the associated morbidity and mortality. We collected data on loco-regional toxicity recorded according to the classification described by Wieberdink et al. [13]. Postoperative mortality was defined as mortality occurring within 30 days of the procedure [14,15]. Whenever two Wieberdink et al. classes were reported together (e.g., 10 cases had class 1 and 2 toxicity), we upgraded patients to the higher grade (grade 2).

**Risk of Bias in Individual Studies**

Previously published reviews have demonstrated that the quality of reporting in published phase II trials and, specifically, in cancer trials is generally poor [16–18]. However, no standardized statement for reporting phase II cancer trials has been created to date. From the EQUATOR Network (www.equator-network.org), we were able to find a paper describing guidelines for standardizing the reporting of phase I and phase II trials in neuro-oncology [16]. The paper included a checklist, many items of which can be generalized to all cancer phase II trials. We adjusted some of these to our area of search and had a final count of 42 ideal criteria, of which we further categorized 16 as essential (Appendix 2). These essential quality criteria include: mention of “trial” or “phase II” in the title, structured abstract, mention of the rationale of the trial in the introduction; report of ethics approval and any conflict of interest; report of consent obtained from eligible patients; detailed report of dosage, techniques, complications, and criteria to define clinical outcomes; appropriate statistical methods section describing at least the hypothesis being tested and sample size calculation; detailed report of patient characteristics (Table I); follow-up of all patients included in the trial; and interpretation of the results in view of the study hypothesis and limitations.

**Summary Measures and Synthesis of Results**

The principal summary measures were proportions of patients analyzed with CR, PR, and LS rates. For our primary end points, we
In order to assess the possibility of reporting pooled estimates for the primary end points, we used a random effects model and calculated the I-squared statistic, introduced by Higgins and Thompson [19], as a measure of the proportion of the overall variation that was attributable to between-study heterogeneity. We considered I-squared values of greater than or equal to 25%, 50%, and 75% as low, moderate, and high heterogeneity, respectively. We used Stata/IC 11.1 (Stata Corp, College Station, TX) for all calculations.

### RESULTS

#### Study Selection

A total of 1,904 articles were identified from our search after removing duplicates (Fig. 1). These were scanned by title and/or abstract, and 1,817 articles were excluded because they were non-relevant to our review topic, or were literature reviews, non-human experiments, non-English publications, or more duplicates. Two authors reviewed 87 full-text articles and a third author resolved disagreements. Of these 87 papers, only 12 were included in the qualitative synthesis; the reasons for exclusion of the others are listed in Figure 1. A summary of the trials involved in the analysis, with corresponding references, is presented in Table I.

### Descriptive Statistics

The total number of patients across all trials was 518. Median ages are presented in Table I; in two trials, only mean age was available. The median rate of loco-regional toxicity per the Wieberdink classification was 3.8%, 45.5%, 17%, 1%, and 0% for levels 1 through 5, respectively. There was one reported case of postoperative mortality [20]. Response assessment modalities were clinical in three trials, pathological in two trials, radiological in two trials, and combined in five trials. Seven of 12 studies included patients with metastatic disease. In eight trials, the chemotherapeutic agent of choice was melphalan combined with TNF-alpha; in four of the eight, interferon gamma was also added to the regime. Three trials tested doxorubicin’s efficacy, and one trial tested doxorubicin alone. In five trials, not all patients received the same treatment regime.

### Primary Outcomes

Overall, of 518 patients across all studies, 408 demonstrated clinical response (CR or PR), and 428 had their treated limb spared. Median rates of CR, PR, and LS were 31%, 53.5%, and 82.5%, respectively. The calculated rates of primary outcomes achieved, with 95% confidence intervals, are illustrated in Table II and Figure 2. Examination of the confidence intervals created around the primary outcomes achieved, with 95% confidence intervals, are illustrated in Table II and Figure 2. Examination of the confidence intervals created around the primary outcomes achieved, with 95% confidence intervals, are illustrated in Table II and Figure 2. Examination of the confidence intervals created around the primary outcomes achieved, with 95% confidence intervals, are illustrated in Table II and Figure 2. Examination of the confidence intervals created around the primary outcomes achieved.
TABLE II. Proportions (Rates) of Patients Exhibiting Complete and Partial Response and Limb Salvage

<table>
<thead>
<tr>
<th>Study</th>
<th>Complete response</th>
<th>Partial response</th>
<th>Limb salvage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Proportion</td>
<td>95% Confidence interval</td>
</tr>
<tr>
<td>Bonvalot et al. [46]</td>
<td>36</td>
<td>0.37</td>
<td>0.27-0.47</td>
</tr>
<tr>
<td>Di Filippo et al. [47]</td>
<td>4</td>
<td>0.28</td>
<td>0.10-0.50</td>
</tr>
<tr>
<td>Di Filippo et al. [48]</td>
<td>25</td>
<td>0.33</td>
<td>0.23-0.45</td>
</tr>
<tr>
<td>Eggermont et al. [8]</td>
<td>54</td>
<td>0.29</td>
<td>0.23-0.36</td>
</tr>
<tr>
<td>Gutman et al. [20]</td>
<td>13</td>
<td>0.37</td>
<td>0.21-0.55</td>
</tr>
<tr>
<td>Lejeune et al. [49]</td>
<td>4</td>
<td>0.21</td>
<td>0.07-0.39</td>
</tr>
<tr>
<td>Lev-Chelouche et al. [50]</td>
<td>1</td>
<td>0.29</td>
<td>0.04-0.64</td>
</tr>
<tr>
<td>Lienard et al. [28]</td>
<td>3</td>
<td>0.67</td>
<td>0.28-0.95</td>
</tr>
<tr>
<td>Rossi et al. [51]</td>
<td>12</td>
<td>0.52</td>
<td>0.33-0.71</td>
</tr>
<tr>
<td>Rossi et al. [52]</td>
<td>2</td>
<td>0.10</td>
<td>0.02-0.24</td>
</tr>
<tr>
<td>Santinami et al. [53]</td>
<td>7</td>
<td>0.67</td>
<td>0.39-0.89</td>
</tr>
<tr>
<td>Wray et al. [30]</td>
<td>1</td>
<td>0.11</td>
<td>0.01-0.27</td>
</tr>
</tbody>
</table>

Fig. 1. PRISMA flow diagram of identification, screening, eligibility, and inclusion of studies.

Journal of Surgical Oncology
Two observations probably explain such extreme heterogeneity in LS outcomes: First, there was no a priori defined time frame in which a case would be counted as a success, that is, that the limb was salvaged. Because it is clinically sensible to assume that LS is a time-dependent outcome, the longer the follow-up time, the higher the rates of failure will be; thus, comparing studies with different follow-up times would result in heterogeneity and pooling them blindly would be unsound. A proposed solution is to estimate a hazard model and to use the per-period hazard as the effect size estimate; however, in order for this to hold, the assumption that the hazard is constant must be met, which is not applicable in this case. Second, Figure 2 clearly shows that the Wray et al. [30] LS rate is different from the general trend of the other trials. Although the median follow-up for that study was 17 months, the investigators followed all patients until they died, and LS status was declared only at that time (whether or not the limb was intact until the time of death). The other studies do not provide details as to when LS was determined; furthermore, we cannot determine if the follow up time was truncated at publication or by metastasis and death.

When looking at trends over time, we found that as the proportions of reported CR decrease, the proportions of PR increase, probably reflecting improved assessment of tumor response to HILP (pathologic and radiologic; Fig. 3).

Quality of Studies and Risk of Bias

Figure 4 presents a summary of the quality assessment of the 12 manuscripts included. None of the trials fulfilled either all ideal or all essential criteria. Only two trials fulfilled more than 75% of essential criteria. Seven trials did not have a statistical method section, and among those trials that did, only one mentioned the determining rule of success and only two reported the calculation of sample size. Finally, only one study mentioned conflict of interest. Despite all of these issues, only two trials addressed their own limitations. This implies that included trials are generally of poor quality, raising suspicions of biased study findings.

DISCUSSION

Despite improved surgical technique, treatment of STS of the extremities can represent a significant challenge. The understanding that amputation does not enhance survival in patients with large (>5 cm) high-grade sarcomas has driven the search for less invasive alternatives. The use of systemic chemotherapy alone in the preoperative context has not translated to meaningful clinical gains, as shown by a large phase 2 trial (EORTC STBSG 62871) [21]. Its combination with preoperative radiotherapy may hold promise [22–24]. In contrast, integration of neoadjuvant radiation therapy in the management of extremity STS currently allows limb-sparing resections to be safely be achieved in as many as 90% of patients [4]. Nevertheless, about 10% need amputation, whether as the only means of eradicating local disease or because a LS attempt would predictably result in a worse functional outcome than amputation. HILP may represent an alternative for these cases.

Although the main concepts of isolated limb perfusion were introduced in the 1950s for melanoma, it was not until the 1980s that treatment regimes were established for the management of extremity sarcomas. Initial results of single-agent case series and reviews were disappointing [25–27]. This lack of efficacy prompted investigators to explore the addition of recombinant TNF-alpha (rTNF-alpha), a powerful antitumor cytotoxic factor [2]. The pioneering work of Lejeune and Lienard showed rTNF-alpha to be a safe and effective agent in a phase II trial of HILP for the management of extremity STS [28]. The use of rTNF-alpha was further supported by the results of a multicenter trial [8], and its use in HILP was approved in European centers [29]. Unfortunately, rTNF-alpha remains unavailable in North America and its use in HILP has not yet been approved by the Food and Drug Administration [30,31].

From our review, we found an overall median response rate for HILP of 81.5%. This is somewhat similar to the results from a recently published systematic review of isolated limb perfusion efficacy and safety in melanoma patients [32]. They reported an overall median response rate of 90% in 1,587 perfusions. They also reported a median rate of regional toxicity of 73.53% for grade 2, 17.1% for
grade 3, and 2.0% for grade 4. Toxic amputation (grade 5 toxicity) was described in 0.65% of treated patients identified in that review. The current study found similar median rates of loco-regional toxicity, with 3.8%, 45.5%, 17%, 1%, and 0% for levels 1 through 5, respectively.

Median rates of CR, PR, and LS in this review were 31%, 53.3%, and 82.5%, respectively. These rates compare well to results of previous trials addressing achievements for local and systemic control in patients with STS of the extremities after combined neoadjuvant chemoradiation therapy [33–36].

In an attempt to summarize the evidence and to see whether enough has been published to proceed to the next level of investigation, randomization, we assessed the quality of phase II trial reports of HILP in STS of the extremities and found that it was generally poor. There is an absolute loss of uniformity in terms of defining potential candidates, technique, dosages, and response outcome, as well as who will assess the response and how and when it will be assessed. This observation of lack of quality in reporting sarcoma clinical trials has also been noted by others. In a recent review by Toulmonde et al., the reporting quality of 72 randomized controlled trials (RCTs) examining the different treatments of sarcoma was assessed. Using the revised Consolidated Standards of Reporting Trials statement, the authors found that the overall quality of sarcoma RCTs reporting key methodological issues was poor [37]. This comes as no surprise, particularly in the field of surgery and when the disease of interest is rare [38,39]. Although RCTs are considered the gold standard for establishing safety and efficacy of an intervention, the overall low frequency of surgical RCTs persists [40,41]. In the context of rare diseases, the question of when to go from innovation to randomization becomes almost impossible to answer. In fact, recent understandings of the shortcomings of surgical RCTs led to greater focus on the complex relation between innovation and practice in surgery. This emphasis occurred based on the understanding that RCTs per se are not the only way to generate valid evidence and that synthesizing and evaluating structured, well-planned, and transparently reported evidence is what modern surgical practice requires [41–43].

One limitation of our review is that it does not include non-English language studies; as most of the trials were conducted in Europe, this raises the possibility of our having missed publications. Moreover, given the fact that all published trials reported HILP as an efficacious procedure, this raises concerns about publication bias. Our analysis was based on published data; we were unable to get updated information from authors contacted. Despite these limitations, our search was exhaustive and comprehensive; to the best of our knowledge, this is the first report of a systematic review of HILP efficacy and quality assessment of published trials. Unlike previously published traditional narrative reviews, systematic reviews aim to
minimize bias in locating, selecting, and interpreting individual studies. Although we were conservative about reporting pooled estimates of primary outcomes (CR, PR, and LS), our decision to not conduct a meta-analysis was based on the extreme heterogeneity of effect across studies. This can be attributed to many factors. First, the chemotherapeutic agent used in the HILP was not standardized across all trials regarding the type of agent used and the exact dosage. Second, although we applied the WHO objective tumor response assessment reporting standards, the method of measuring the response (clinical, radiological, or pathological) was not the same across all trials. Third, we had to account for the fact that for the determination of LS, follow-up time was not standardized across studies; therefore, reporting a pooled effect estimate would be premature.

CONCLUSION

Sarcoma is a rare disease, and HILP is a complex procedure performed in a limited number of centers, possibly explaining why we did not identify many well-designed trials. However, the review of the current literature is disappointing in that it did not allow us to identify a single publication of good enough quality to justify a phase III trial comparing this promising intervention to alternative, more traditional approaches. Since RCTs are not feasible at the moment, well-designed, national, multi-institutional phase II trials are warranted to examine the efficacy of HILP in the treatment of STS of the extremities. These trials should be of high quality and transparently reported. In the absence of that, there is a need for greater standardization of HILP protocol details and chemotherapeutic regimens used, validation and harmonization of outcomes, and centralized reporting for adequate follow-up. The communication and interaction of centers through a structured, organized network helped with effective decision making and standardization of methodological consensus in the treatment of peritoneal surface malignancies with cytoreductive surgery and intraperitoneal chemotherapy [44,45]. This approach can serve as a model in standardizing HILP selection criteria and assessment methods in STS patients.

ACKNOWLEDGMENTS

We thank Barbara Every, ELS, of BioMedical Editor, for editing the manuscript. We also thank Lorie Kloda, MLIS, AHIP; Associate Librarian at McGill University for help with structuring of the search strategy and the review protocol.

REFERENCES

27. Schwarzbach M, Lehnten T, Willeke F, et al.: Results of isolated hyperthermic extremity perfusion in soft tissue sarcomas within the
8 Trabulsi et al.