Critical Evaluation of Extracorporeal Photopheresis for Treatment of Chronic Graft-versus-host Disease

Paul J. Martin, M.D.
Fred Hutchinson Cancer Research Center
University of Washington

Disclosure: No relevant disclosures for this presentation
Note: Chronic GVHD is not an approved indication for ECP
The Story Line

• How I got here—The backstory on a different treatment
• A critical evaluation of ECP studies
• Where we go from here
<table>
<thead>
<tr>
<th>Study</th>
<th>Indication</th>
<th>N</th>
<th>CR (%)</th>
<th>PR</th>
<th>CR + PR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Secondary</td>
<td>26</td>
<td>2</td>
<td>10</td>
<td>12 (46)</td>
</tr>
<tr>
<td>2</td>
<td>Secondary</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>5 (100)</td>
</tr>
<tr>
<td>3</td>
<td>Secondary</td>
<td>21</td>
<td>5</td>
<td>8</td>
<td>13 (62)</td>
</tr>
<tr>
<td>4</td>
<td>Secondary</td>
<td>15</td>
<td>2</td>
<td>7</td>
<td>9 (60)</td>
</tr>
<tr>
<td>5</td>
<td>Secondary</td>
<td>11</td>
<td>3</td>
<td>5</td>
<td>8 (73)</td>
</tr>
<tr>
<td>6</td>
<td>Secondary</td>
<td>13</td>
<td>1</td>
<td>9</td>
<td>10 (77)</td>
</tr>
<tr>
<td>7</td>
<td>Secondary</td>
<td>13</td>
<td>4</td>
<td>5</td>
<td>9 (69)</td>
</tr>
<tr>
<td>8</td>
<td>Secondary</td>
<td>24</td>
<td>5</td>
<td>13</td>
<td>18 (75)</td>
</tr>
<tr>
<td>Total</td>
<td>Secondary</td>
<td>128</td>
<td>24 (19)</td>
<td>60</td>
<td>84 (66)</td>
</tr>
<tr>
<td>9</td>
<td>Primary</td>
<td>10</td>
<td>7 (70)</td>
<td>2</td>
<td>9 (90)</td>
</tr>
</tbody>
</table>
Agents Used for Treatment of Chronic GVHD

- Corticosteroids
- Cyclosporine
- Topical corticosteroids
- Tacrolimus
- Mycophenolate mofetil
- Corticosteroid mouthwash
- PUVA
- Thalidomide
- Extracorporeal photopheresis
- Anti-thymocyte globulin
- Ursodeoxycholic acid (Actigal)
- Daclizumab (Zenapax)
- Azathioprine (Imuran)
- Hydroxychloroquine (Plaquenil)
- Infliximab (Remicade)

Percentage of Respondents

- Used, great success
- Used, some success
- Used, no success
- Not used/missing

Lee et al. BBMT 2002;8:32-39
Initial Treatment of High-Risk Chronic GVHD*

<table>
<thead>
<tr>
<th>Recommended treatment</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add MMF</td>
<td>54</td>
</tr>
<tr>
<td>Add psoralen and UVA</td>
<td>17</td>
</tr>
<tr>
<td>Add sirolimus</td>
<td>15</td>
</tr>
<tr>
<td>Add rituxan</td>
<td>10</td>
</tr>
<tr>
<td>Continue steroid treatment</td>
<td>5</td>
</tr>
</tbody>
</table>

*Pavletic and Vogelsang: ASBMT eNEWS Survey, 2008
Controlled Trial of MMF for First-Line Treatment

• Prednisone plus MMF versus prednisone plus placebo
• Double-blind
• Primary endpoint, survival to resolution of chronic GVHD and withdrawal of systemic immunosuppressive treatment without recurrent malignancy
The graph shows the percent treatment success over months from randomization. The solid line represents MMF, and the dashed line represents Placebo. The vertical dashed line at 24 months indicates a significant difference in success rates between the two groups. By 36 months, the MMF group shows a higher success rate compared to the Placebo group.
Response at 3 Months*

*Excludes secondary therapy, relapse or death before 3 months
Response at 3 Months*

*Includes failures before 3 months: 24 in MMF arm and 11 in placebo arm

P = .003
MMF ——— Placebo

Survival

Non-relapse Mortality

Relapse-free Survival

Recurrent Malignancy
Potential Reasons for Lack of Success

- Imbalanced risk factors
- Patient non-compliance
- MMF dose not optimal
The Most Likely Reason for Lack of Success

- Imbalanced risk factors
- Patient non-compliance
- MMF dose not optimal
- MMF not effective for this indication
Why Were the Prior Studies of MMF So Misleading?

- Post hoc, ergo propter hoc fallacy?
- Poor data quality?
- Poor study design?
When Can We Trust “Post hoc, ergo propter hoc” Reasoning?

- Well-defined eligibility criteria
- Stable or worsening trajectory of disease before enrollment
- No change of treatment within 1 month before enrollment
- No other new treatment before endpoint assessment
- Consistent treatment regimen
The Importance of High-quality Data

- Well-defined eligibility criteria
- Stable or worsening trajectory of disease before enrollment
- No change of treatment within 1 month before enrollment
- No other new treatment before endpoint assessment
- Consistent treatment regimen
- Use of case-report form to collect data
- Use of objective organ-severity measures
- Well-defined overall response criteria
Critical Design Characteristics

• Well-defined eligibility criteria
• Stable or worsening trajectory of disease before enrollment
• No change of treatment within 1 month before enrollment
• No other new treatment before endpoint assessment
• Consistent treatment regimen
• Use of case-report form to collect data
• Use of objective organ-severity measures
• Well-defined overall response criteria
• Response measured at an appropriate time-point
• Primary endpoint reflects clinical benefit
• Pre-specified statistical design with an appropriate benchmark
• Analysis accounts for baseline risk factors
Ten Quality Indicators

- Defined eligibility criteria
- Demonstration that cohort is representative (or not)
- Consistent treatment regimen
- Objective criteria for organ response
- Specified overall response criteria
- Specified time for assessment of response
- Concomitant treatment taken into account
- Well established control benchmark
- Pre-specified statistical hypothesis
- Statement of statistical power
# Quality of Prior Studies of MMF

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Study Identification and Type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligibility criteria</td>
<td>R</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>P</td>
<td>R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Representative cohort</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistent treatment regimen</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective response criteria</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall response criteria</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of assessment</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Concomitant treatment</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historical benchmark</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Statistical hypothesis</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Statistical power</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions Stated in Prior Studies

• Retrospective studies
  – “is a potentially useful”
  – “suggest.....can be effective for some”
  – “would appear to be effective and safe”
  – “can be efficiently and safely used”
  – “suggest....is an active, well-tolerated agent”

• Prospective studies
  – “suggest....has a beneficial effect”
  – “suggest....may have a role to play”
  – “suggest....may be a useful treatment”
  – “can be used effectively”

Conclusion: Abstract: n = 7; Discussion: n = 2
Call for additional studies: Abstract; n = 3; Discussion: n = 6
A More Appropriate Conclusion

• Our results demonstrate the feasibility of using MMF to treat chronic GVHD. The true merits of using MMF for this indication can be evaluated only in a prospective controlled trial.
Lessons Learned

• Conclusions from retrospective studies and phase II studies should be stated more cautiously

• Small retrospective studies have very limited value for assessing results of a new treatment

• The distinction between retrospective studies and prospective studies is important

• Many phase II studies fall far short of the ideal

• Repetition of phase II studies does not advance the field

• Need to move more quickly from phase II studies to definitive phase III studies, especially in testing agents already approved for other indications
Quality of 8 Prospective ECP Studies 1998 – 2010*

Only One Controlled Study of ECP*

• Design
  – Randomized ECP vs. continued prednisone, crossover at 12 weeks

• Strengths
  – Robust eligibility criteria
  – Selection bias minimized
  – Objective measurement of skin involvement

• Weaknesses
  – No categorical definition of clinically significant response
  – No assessment of overall response
  – Short follow-up before crossover (12 weeks)
  – No statistical hypothesis

• Results
  – No difference in skin score between arms at 12 weeks

*Flowers et al. Blood 2008;112: 2667-2674
Author’s Conclusions and My Critique

• “These results suggest that ECP may have a steroid-sparing effect in the treatment of chronic GVHD”

• Favorable conclusion was based entirely on open-label results
  – Physicians’ assessments of skin response
  – Physicians’ decisions to decrease steroid dose
  – Composite endpoint of objective skin response and steroid dose
## ECP Publications since 2010

<table>
<thead>
<tr>
<th>Type of Publication</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>6</td>
</tr>
<tr>
<td>Case report</td>
<td>7</td>
</tr>
<tr>
<td>Laboratory correlates</td>
<td>4</td>
</tr>
<tr>
<td>Retrospective study</td>
<td>7</td>
</tr>
<tr>
<td>Prospective study</td>
<td>2</td>
</tr>
<tr>
<td>Review</td>
<td>18</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>44</strong></td>
</tr>
</tbody>
</table>
## ECP Publications since 2010

<table>
<thead>
<tr>
<th>Type of Publication</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>6</td>
</tr>
<tr>
<td>Case report</td>
<td>7</td>
</tr>
<tr>
<td>Laboratory correlates</td>
<td>4</td>
</tr>
<tr>
<td>Retrospective study</td>
<td>7</td>
</tr>
<tr>
<td>Prospective study</td>
<td>2</td>
</tr>
<tr>
<td>Review</td>
<td>18</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>44</td>
</tr>
</tbody>
</table>
Follow-up after ECP Crossover*

• Design
  – Comparison before and after crossover (N = 29)

• Strengths
  – Controlled design with before and after comparison
  – Objective serial measurements of skin involvement

• Weaknesses
  – No categorical definition of clinically significant skin response
  – Absolute change in skin score not reported
  – No statistical analysis of changes in skin score
  – No assessment of overall response

• Results
  – Progressive improvement in skin score beginning after 12 weeks
  – Absolute change in skin score at 24 weeks is not likely to be clinically significant

*Greinix et al. BBMT 2011;17: 1782-2011
Where Do We Go From Here?

- Response does not equate to clinical benefit
- Assess at later time points: chronic GVHD is a **chronic** disease
- Use a response definition associated with shortened time to end of treatment and improved survival
A “Response” that Could Indicate Clinical Benefit

- Specified time interval for endpoint assessment
  - 1 year for first-line treatment
  - 6 months for second or subsequent-line treatment

- Improvement in disease manifestations
  - Complete resolution of all disease manifestations, or
  - Clinically significant improvement in at least 1 disease manifestation, including the most bothersome symptom, without worsening in any other manifestations

- No new systemic treatment after enrollment
  - May also include limits on new topical treatment, escalated steroid dosing during taper, and upper steroid dose at endpoint assessment

- Competing risks counted as failure
  - Death before endpoint assessment
  - Recurrent malignancy before endpoint assessment
A Possibly More Realistic Endpoint

• Specified time interval for endpoint assessment
  – 1 year for first-line treatment
  – 6 months for second or subsequent-line treatment

• Prevention of “Progressive Impairment”
  – Emergence of an enduring chronic GVHD-related health state that
    threatens or compromises a patient’s physical well-being or function
    in ways that cannot be easily reversed

• No new systemic treatment after enrollment
  – May also include limits on new topical treatment, escalated steroid
    dosing during taper, and upper steroid dose at endpoint assessment

• Competing risks counted as failure
  – Death before endpoint assessment
  – Recurrent malignancy before endpoint assessment
Possible Study Designs

• First-line treatment is amenable to single-arm designs
  – Outcome compared to well-established historical benchmark

• Treatment beyond first-line requires controlled designs
  – Outcomes compared between arms
  – New systemic treatment before assessment counted as failure in the investigational arm
  – New systemic treatment may be allowed before assessment in the control arm
  – Death and recurrent malignancy before assessment counted as failure in both arms
Mapping a Path Toward Genuine Progress

Biology of Blood and Marrow Transplantation

Report


Paul J. Martin 1,*, Stephanie J. Lee 1, Donna Przepiorka 2, Mary M. Horowitz 3, John Koreth 4, Georgia B. Vogelsang 5, Irwin Walker 6, Paul A. Carpenter 1, Linda M. Griffith 7, Gorgun Akpek 8, Mohamad Mohty 9, Daniel Wolff 10, Steven Z. Pavletic 11, Corey S. Cutler 4

1 Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington
2 Center for Drug Evaluation Research, US Food and Drug Administration, Silver Spring, Maryland
3 Division of Hematology and Oncology, Medical College of Wisconsin, Milwaukee, Wisconsin
4 Division of Hematologic Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts
5 Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland
6 Department of Medicine, McMaster University, Hamilton, Ontario, Canada
7 Division of Allergy, Immunology and Transplantation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland
8 Stem Cell Transplant Program, Banner MD Anderson Cancer Center, Gilbert, Arizona
9 Clinical Hematology and Cellular Therapy Department, Hôpital Saint-Antoine, University Pierre & Marie Curie, INSERM U938, Paris, France
10 Department of Internal Medicine III, University of Regensburg, Regensburg, Germany
11 Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

*BBMT 2015;21:1343-1359
Special Acknowledgments

- Yoshi Inamoto
- Stephanie Lee
- Shawn Chai
- Barry Storer
- NIH Consensus Conference on Chronic GVHD
- U.S. FDA, Division of Hematology Products
- National Cancer Institute: CA18029 and CA98906