ASK THE QUESTION

**Question:** In newborns with hyperbilirubinemia, what conventional phototherapy delivery method (i.e., single, double, triple) is the most effective for reducing bilirubin levels?

SEARCH FOR EVIDENCE

**Databases:** PubMed, Scopus, CINAHL

**PubMed search strategy:** ("Phototherapy"[Mesh] OR phototherap* OR photoradiation OR "light therapy") AND ("Hyperbilirubinemia, Neonatal"[Mesh] OR hyperbilirubinemia OR jaundice OR icterus) with and without ("Comparative Effectiveness Research"[Mesh] OR comparative OR effectiveness OR efficacy)

**Filters:** Humans, English

CRITICALLY ANALYZE THE EVIDENCE

There were five studies found addressing the effectiveness of different conventional (i.e., florescent, halogen) phototherapy delivery methods in newborns with hyperbilirubinemia. This included one systematic review (Mills et al., 2001), three randomized controlled trials (Abd Hamid et al., 2013; Boonyarittipong et al., 2008; Naderi et al., 2009), and one descriptive secondary analysis of RCT data (Morris et al., 2013).

A systematic review by Mills et al. (2001) established a role for fiber optic phototherapy (i.e., BiliBlankets, Wallaby phototherapy) in treating newborns with hyperbilirubinemia. While the percentage change in serum bilirubin levels at 24 and 48 hours was significantly lower with BiliBlankets alone compared to conventional therapy, there was a trend toward an increased percentage change in serum bilirubin levels at both time points for combination therapy (florescent light bank + fiber optics) over conventional therapy (WMD -3.2%, 95%CI -17.2 to 10.8; 1 study and WMD -9.2%, 95%CI -25.02 to 6.62; 1 study, respectively).

Abd Hamid et al. (2013) randomly assigned newborns to receive phototherapy using a single bank of florescent lights plus a silver-colored reflective curtain (n=80) or a double bank of florescent lights (n=80). They found that there was no significant difference in the mean decrease in serum bilirubin after 4 h of phototherapy (single + curtain: 22.70 ±27.70 umol/L and double: 22.53±28.55 umol/L, p = 0.97). Additionally, only four babies developed rebound in serum
bilirubin levels needed to restart phototherapy again (single = 2, double = 2). Boonyarittipong et al. (2008) randomly assigned healthy, full-term infants with non-hemolytic hyperbilirubinemia to single (n=30; above) or double (n=30; above and below) phototherapy using florescent light banks. They found a **steeper decline in bilirubin levels for the double surface group compared to the single surface group (5.4±2.0 mg/dl vs. 3.5±1.7; p < 0.001)**. However, after 24 hours of phototherapy the mean serum bilirubin levels were not significantly different between the groups (p=0.05). Naderi et al. (2009) randomized healthy, term newborns with indirect hyperbilirubinemia to double (n=20; above and lengthwise) or triple (n=20; above, lengthwise, widthwise) phototherapy using florescent light banks. They found no significant difference in the rate of bilirubin decline within the first 8 (p=0.590), 16 (p=0.760) and 24 (p=0.370) hours of double or triple phototherapy. Mean LOS (hr) was also not significantly different between the two groups.

Finally, Morris et al. (2013) evaluated the efficacy of different phototherapy devices (florescent banks, fiber optic blankets, halogen spots, LED) and the outcomes of 1392 extremely premature infants. The **adjusted absolute decrease in total serum bilirubin for LEDs was significantly greater than for halogen spots (p=0.0038), florescent banks (p<0.0001) and fiber optic blankets (p<0.001).** Additionally, the study found that at 18-22 months there was an increased risk of death or a mental development index < 85 when florescent banks were used for phototherapy compared to LEDs, halogen spots, and fiber optic blankets.

### PICO Question

In newborns with hyperbilirubinemia, what conventional phototherapy delivery method (i.e., single, double, triple) is the most effective for reducing bilirubin levels?

<table>
<thead>
<tr>
<th>Author/Date/ Journal</th>
<th>Purpose of Study</th>
<th>Study Design</th>
<th>Sample&amp; Setting</th>
<th>Outcomes</th>
<th>Design Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mills et al., 2001, Cochrane Database of Systematic Reviews</td>
<td>To evaluate the efficacy of fiber optic phototherapy</td>
<td>Systematic review &amp; meta-analysis</td>
<td>24 RCTs and quasi-randomized studies -newborns up to 28 days old</td>
<td>Fiber optic vs Conventional (24 studies): - percentage change in serum bilirubin [SBR] after 24 hr (WMD 3.59%, 95% CI 1.27-5.92; 7 studies) and 48 hr (WMD 10.79%, 95% CI 8.33-13.26; 6 studies) of treatment was greater in the conventional phototherapy group -use of exchange transfusion in the fiber optic group was increased but did not reach statistical significance (RR 1.62, 95% CI 0.38-6.93; 4 studies) - use of additional phototherapy in the fiber optic group was significantly increased (RR 1.68, 95% CI 1.18-2.38; 10 studies) - use of repeat phototherapy for rebound jaundice was not different between the groups (RR 2.33, 95% CI 0.92-5.91; 6 studies)</td>
<td>Study Limitations = None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BiliBlanket vs Conventional (13 studies):</td>
<td></td>
</tr>
</tbody>
</table>

**GRADE CRITERIA**

(See Appendix A)

Lower Quality Rating if:

- High risk of bias
  - (When design limitations for one or more criteria impact the quality of studies sufficiently enough to lower confidence in the estimate of effect)

- Studies inconsistent
  - (When there are differences in the direction of the effect, populations, interventions or outcomes between studies)

- Studies are indirect
  - (Your PICO question is quite different from the available evidence in regard to population, intervention, comparison, or outcome)

- Studies are imprecise
  - (When studies include few patients and few
<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Participants</th>
<th>Intervention</th>
<th>Outcome Measures</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abd Hamid et al., 2013, Journal of Pediatrics and Child Health</td>
<td>RCT</td>
<td>160 babies in a single NICU in Malaysia (2010-11)</td>
<td>Single phototherapy: (n=80) 1 unit + reflective curtains (silver-colored reflective cloth on 3/4 sides)</td>
<td>Intent to treat result showed that there was no significant difference between both groups in the mean decrease in serum bilirubin after 4 h of phototherapy (single: 22.70 ±27.70 umol/L and double: 22.53±28.55 umol/L, p = 0.97)</td>
<td>None</td>
</tr>
</tbody>
</table>

- percentage change in SBR after 24 hr (WMD 3.51%, 95%CI 0.76-6.25; 5 studies) and 48 hr (WMD 8.27%, 95%CI 4.62-11.92; 4 studies) was significantly less in the BiliBlanket group - use of additional phototherapy was significantly increased in the BiliBlanket group (RR 1.57, 95%CI 1.10-2.24; 9 studies) - use of exchange transfusion (RR 1.62, 95%CI 0.38-6.93; 4 studies) and repeat phototherapy for rebound jaundice (RR 1.72, 95%CI 0.70-4.27; 5 studies) were not significantly different

Combination vs Conventional (6 studies):
- trend to greater percentage change in SBR at 24 hr (WMD -3.2%, 95%CI -17.2 to 10.8; 1 study) and 48 hr (WMD -9.2%, 95%CI -25.02 to 6.62; 1 study) in the combination group
- trend to less use of exchange transfusion (RR 0.24, 95%CI 0.01, 4.72, 1 study) and additional phototherapy (RR 0.11, 95%CI 0.01, 2.02; 2 studies) in the combination group

Events and thus have wide confidence intervals and the results are uncertain

- Publication Bias (e.g. pharmaceutical company sponsors study on effectiveness of drug)
- Increase Quality Rating if:
  - Large effect (When the relative risk of association between two factors is large or very large)
  - Dose response (When the dose-response relationship increases the confidence than an effect is real and substantial)
  - Plausible confounders (When plausible residual confounding is directly impacting the magnitude of effect)

Level of evidence for studies as a whole:
- High
- Moderate
- Low
- Very Low
| Boonyarittipong et al., 2008, Journal of the Medical Association of Thailand | To compare the efficacy and adverse effects of double-surface and single-surface intensive phototherapy in term newborn infants with hyperbilirubinemia | RCT | 60 healthy, full-term infants with non-hemolytic hyperbilirubinemia (13.0-19.9 mg/dl) at a single hospital in Thailand (2006-07) - birthweight > 2500g | After 24 hours of phototherapy, the mean serum bilirubin in the single surface group was 11.3±2.1mg/dL, while in the double surface group it was 10.3±1.9mg/dL, the difference was not statistically significant (p = 0.05) There was a steeper decline in bilirubin levels for the double surface group compared to the single surface group (5.4±2.0 mg/dL vs. 3.5±1.7; p < 0.001) - cumulative declines after 48 hours were also significantly steeper in the double surface group (8.4±2.1 mg/dL vs. 6.5±2.3 mg/dL; p = 0.001) Body temperature 24h after phototherapy was higher in the double surface group (+0.2°C; p=0.003) and defecation 24 and 48 hours after phototherapy were both higher in the single surface group (p=0.04 and p=0.001, respectively) | Study Limitations = None  
RCT & Quasi-Experimental Studies  
× Insufficient sample size  
× Lack of randomization  
× Lack of blinding  
× Stopped early for benefit  
× Lack of allocation concealment  
× Selective reporting of measures  
× Large losses to F/U |
<table>
<thead>
<tr>
<th>Study Authors</th>
<th>Research Question</th>
<th>Study Design</th>
<th>Participants</th>
<th>Outcomes</th>
<th>Study Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naderi et al., 2009, Pediatrics and Neonatology</td>
<td>To compare the efficacy and length of hospital stay of double and triple phototherapy in newborns with indirect hyperbilirubinemia</td>
<td>RCT</td>
<td>40 healthy, term newborns with indirect hyperbilirubinemia (&gt; 12 and &gt; 15 mg/dL in the 2nd and 3rd day) at a single hospital in Iran (2006) - birthweight &gt; 2500g</td>
<td>There was no significant difference in the rate of bilirubin decline between the two groups within the first 8 (double: 13.7 ± 1.8 mg/dL vs triple: 14.1 ± 1.7 mg/dL; p=0.590), 16 (double: 12.2 ± 1.7 mg/dL vs triple: 12.4 ± 1.7 mg/dL; p=0.760), and 24 (double: 10.3 ± 2 mg/dL vs triple: 10.9 ± 1 mg/dL; p=0.370) hours</td>
<td>Mean LOS (hr) was not significantly different between the 2 groups (double: 34.6 ± 16.5 vs triple: 41.5 ± 17.7; p=0.211) Analysis of complications showed no difference between triple and double groups</td>
</tr>
<tr>
<td>Morris et al., 2013, Journal of Perinatology</td>
<td>To evaluate the efficacy of different phototherapy devices and the outcomes of extremely premature infants</td>
<td>Secondary analysis of RCT data - National Institute of Child Health and Human Development (NICHD) Neonatal Research Network</td>
<td>1392 premature, extremely low birthweight babies (501-1000 g) treated with at least one phototherapy device over a 2-week period</td>
<td>The adjusted absolute decrease in total serum bilirubin [TSB] for LEDs (-2.2 mg dL⁻¹) was significantly greater than that for Spots (-1.7 mg dL⁻¹; p=0.0038), Banks (-1.3 mg dL⁻¹; p&lt;0.0001) and Blankets (-0.8 mg dL⁻¹; p&lt;0.001). Infants with &gt;1.5 mg dL⁻¹ decrease in TSB in the first 24 hr: - LEDs were more likely to achieve this decrease than either Blankets or Banks (p=0.0006 and p=0.0066, respectively) - Spots were more likely to achieve this decrease than Blankets (p=0.0070). LEDs had the shortest duration of phototherapy, significantly less than Blankets or Banks.</td>
<td>Study Limitations = None</td>
</tr>
</tbody>
</table>
-Blankets (27+9 µW cm⁻² nm)
-Spots (24+9 µW cm⁻² nm)
-LED (24+6 µW cm⁻² nm)

(p= 0.0011 and p= 0.0014, respectively), but not different from Spots (p= 0.18)
For death or mental development index <85, Banks showed an increased risk compared with the other devices at 18-22 months:
-LEDs (RR 1.16, 95% CI 1.02-1.33)
-Spots (RR 1.26 95% CI 1.08-1.48)
-Blankets (RR 1.43 95% CI 1.12-1.85)

REFERENCES

Appendix A: GRADE criteria for rating a body of evidence on an intervention
Developed by the GRADE Working Group

Grades and interpretations:
High: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low: Any estimate of effect is very uncertain.

Type of evidence and starting level
Randomized trial–high
Observational study–low
Any other evidence–very low

Criteria for increasing or decreasing level
Reductions
Study quality has serious (−1) or very serious (−2) problems
Important inconsistency in evidence (−1)
Directness is somewhat (−1) or seriously (−2) uncertain
Sparse or imprecise data (−1)
Reporting bias highly probable (−1)

Increases
Evidence of association† strong (+1) or very strong (+2)
Dose-response gradient evident (+1)
All plausible confounders would reduce the effect (+1)

†Strong association defined as significant relative risk (RR 2-5 or 0.5-0.2) based on consistent evidence from two or more studies with no plausible confounders;
Very strong association defined as significant relative risk (RR >5 or <0.2) based on direct evidence with no threats to validity

Appendix B. Trustworthy Guideline rating scale
The University of Pennsylvania’s Center for Evidence-Based Practice Trustworthy Guideline rating scale is based on the Institute of Medicine’s “Standards for Developing Trustworthy Clinical Practice Guidelines” (IOM), as well as a review of the AGREE Enterprise and Guidelines International Network domains.

The purpose of this scale is to focus on the weaknesses of a guideline that may reduce the trust a clinical user can have in the guideline, and distinguish weaknesses in documentation (e.g. guideline does not have a documented updating process) from weaknesses in the guidance itself (e.g. recommendations are outdated). Current quality scales like AGREE emphasize documentation. They are important checklists for developers of new guidelines, but are less useful for grading existing guidelines. These scales also are harder for clinicians and other persons who are not methodology experts to apply, and their length discourages their use outside formal technology assessment reports. This new scale is brief, balanced, and easy and consistent to apply.

We do not attempt to convert the results of this assessment into a numeric score. Instead we present a table listing the guidelines and how they are rated on each standard. This facilitates qualitative understanding by the reader, who can see for what areas the guideline base as a whole is weak or strong as well as which guidelines are weaker or stronger.

### 1. Transparency

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Guideline development methods are fully disclosed.</td>
</tr>
<tr>
<td>B</td>
<td>Guideline development methods are partially disclosed.</td>
</tr>
<tr>
<td>C</td>
<td>Guideline development methods are not disclosed.</td>
</tr>
</tbody>
</table>

The grader must refer to any cited methods supplements or other supporting material when evaluating the guideline. Methods should include:

- Who wrote the initial draft
- How the committee voted on or otherwise approved recommendations

Evidence review, external review and methods used for updating are not addressed in this standard.

### 2. Conflict of interest

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Funding of the guideline project is disclosed, disclosures are made for each individual panelist, and financial or other conflicts do not apply to key authors of the guideline or to more than 1 in 10 panel members.</td>
</tr>
<tr>
<td>B</td>
<td>Guideline states that there were no conflicts (or fewer than 1 in 10 panel members), but does not disclose funding source.</td>
</tr>
<tr>
<td>C</td>
<td>Lead author, senior author, or guideline panel members (at least 1 in 10) have conflict of interest, or guideline project was funded by industry sponsor with no assurance of independence.</td>
</tr>
<tr>
<td>NR</td>
<td>Guideline does not report on potential conflict of interests.</td>
</tr>
</tbody>
</table>

For purposes of this checklist, conflicts of interest include employment by, consulting for, or holding stock in companies doing business in fields affected by the guideline, as well as related financial conflicts. This definition should not be considered exclusive. As much as anything, this is a surrogate marker for thorough reporting, since it may be assumed that guideline projects are funded by the sponsoring organization and many authors think it unnecessary to report a non-conflict.

### 3. Guideline development group

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Guideline development group includes 1) methodological experts and clinicians and 2) representatives of multiple specialties.</td>
</tr>
<tr>
<td>B</td>
<td>Guideline development group includes one of the above, but not both.</td>
</tr>
<tr>
<td>C</td>
<td>Guideline developers all from one specialty or organization, and no methodologists.</td>
</tr>
</tbody>
</table>
NR | Affiliations of guideline developers not reported
---|---
The purpose of this standard is to ensure that supporters of competing procedures, or clinicians with no vested interest in utilization of one procedure or another, are involved in development of the guideline. Both AGREE II and IOM call for patient or public involvement: very few guideline panels have done so to date, so this is not necessary for guidelines to be rated A. Involvement of methodologists or HTA specialists in the systematic review is sufficient involvement in the guideline development group for our purposes. In the absence of any description of the guideline group, assume the named authors are the guideline group.

### 4. Systematic review

<table>
<thead>
<tr>
<th>A</th>
<th>Guideline includes a systematic review of the evidence or links to a current review.</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Guideline is based on a review which may or may not meet systematic review criteria.</td>
</tr>
<tr>
<td>C</td>
<td>Guideline is not based on a review of the evidence.</td>
</tr>
</tbody>
</table>

In order to qualify as a systematic review, the review must do all of the following:
- Describe itself as systematic or report search strategies using multiple databases
- Define the scope of the review (including key questions and the applicable population)
- Either include quantitative or qualitative synthesis of the data or explain why it is not indicated

Note: this element does not address the quality of the systematic review: simply whether or not it exists. Concerns about quality or bias of the review will be discussed in text, where the analyst will explain whether the weaknesses of the review weaken the validity or reliability of the guideline.

Note: a guideline may be rated B on this domain even if the review on which it is based is not available to us. This potential weakness of the guideline should be discussed in text of the report.

### 5. Grading the supporting evidence

<table>
<thead>
<tr>
<th>A</th>
<th>Specific supporting evidence (or lack thereof) for each recommendation is cited and graded</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Specific supporting evidence (or lack thereof) for each recommendation is cited but the recommendation is not graded.</td>
</tr>
<tr>
<td>C</td>
<td>Recommendations are not supported by specific evidence.</td>
</tr>
</tbody>
</table>

To score a B on this domain there should be specific citations to evidence tables or individual references for each relevant recommendation in the guideline, or an indication that no evidence was available. Any standardized grading system is acceptable for purposes of this rating. If a guideline reports that there is no evidence available despite a thorough literature search, it may be scored B on this domain, or even A if evidence for other recommendations is cited and graded.

### 6. Recommendations

<table>
<thead>
<tr>
<th>A</th>
<th>Considerations for each recommendation are documented (i.e. benefits and harms of a particular action, and/or strength of the evidence); and recommendations are presented in an actionable form.</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Either one or the other of the above criteria is met.</td>
</tr>
<tr>
<td>C</td>
<td>Neither of the above criteria are met</td>
</tr>
</tbody>
</table>

In order to be actionable, the guideline should specify the specific population to which the guideline applies, the specific intervention in question, and the circumstances under which it should be carried out (or not carried out). The language used in the recommendations should also be consistent with the strength of the recommendation (e.g. directive and active language like “should” or “should not” for strong recommendations, and passive language like “consider” for weak recommendations). A figure or algorithm is considered actionable as long as it is complete enough to incorporate all the applicable patients and interventions. Please see the forthcoming NICE manual (24) for a good discussion of actionability in guidelines.
7. External review

| A | Guideline was made available to external groups for review. |
| B | Guideline was reviewed by members of the sponsoring body only. |
| C | Guideline was not externally reviewed. |
| NR | No external review process is described. |

8. Updating and currency of guideline

| A | Guideline is current and an expiration date or update process is specified. |
| B | Guideline is current but no expiration date or update process is specified. |
| C | Guideline is outdated. |

A guideline is considered current if it is within the developers’ stated validity period, or if no period or expiration data is stated, the guideline was published in the past three years (NOTE: the specific period may be changed at the analyst’s discretion, based on whether the technology is mature and whether there is a significant amount of recent evidence). A guideline must address new evidence when it is updated. A guideline which is simply re-endorsed by the panel without searching for new evidence must be considered outdated.