Utility of Urine Myoglobin for the Prediction of Acute Renal Failure in Patients with Suspected Rhabdomyolysis: A Systematic Review

Karina Rodríguez-Capote, 1 Cynthia M. Balion, 1, 2* Stephen A. Hill, 1, 2 Richard Cleve, 1 Lufang Yang, 1 and Adell El Sharif 1

BACKGROUND: Urine myoglobin continues to be used as a marker of rhabdomyolysis, particularly to assess risk of developing acute renal failure and evaluate treatment success. We sought to determine the predictive validity of urine myoglobin (uMb) for acute renal failure (ARF) in patients with suspected rhabdomyolysis.

METHODS: We performed a broad systemic review of the literature from January 1980 to December 2006 using the search terms myoglobin$ AND (renal OR ARF OR kidney). Only primary studies published in English where uMb measurement was related to ARF were included.

RESULTS: Of 1602 studies screened, 52 met all selection criteria. The studies covered a wide spectrum of etiologies for rhabdomyolysis, dissimilar diagnostic criteria for ARF and rhabdomyolysis, and various methods of uMb measurement and were mostly case series (n = 32). There was poor reporting on the uMb method, and 17 studies failed to provide any information about the method. The reporting of clinical criteria for ARF with respect to timing, description, performance, and interpretation also lacked adequate detail for replication. Eight studies (total 295 patients) had data for 2-by-2 tables. Sensitivity of the uMb test was 100% in 5 of the 8 studies, specificity varied widely (15% to 88%), and CIs around these measures were high. Pooling of data was not possible because of study heterogeneity.

CONCLUSIONS: There is inadequate evidence evaluating the use of uMb as a predictor of ARF in patients with suspected rhabdomyolysis.

© 2009 American Association for Clinical Chemistry

Rhabdomyolysis is a clinical and laboratory syndrome resulting from the breakdown of muscle fibers with release of muscle cell contents into plasma. Myoglobin (Mb), a 17-kDa single-chain oxygen-carrying hemoprotein, appears in the circulation within a few hours of skeletal or cardiac muscle damage and is rapidly filtered by the glomeruli and reabsorbed by the proximal tubules where it is catabolized (1–3). When the filtered load exceeds the reabsorptive capacity of the tubule, myoglobin spills over into the urine, coloring it red (1–5). Rhabdomyolysis can occur by direct muscle injury, ischemia, excessive muscular activity, drugs and toxins, infection, inflammatory myopathies, electrolyte and endocrine/metabolic disorders, hereditary disorders, and temperature extremes. The complication of rhabdomyolysis is acute renal failure (ARF), and depending on the severity and duration of the renal dysfunction, it can lead to chronic renal failure, damage to the heart or nervous system, and death (1, 2, 4–7).

The pathophysiology of myoglobin-induced ARF has not been fully elucidated, but 3 major mechanisms have been proposed, the combination of which contributes to the overall renal damage. One mechanism is physical obstruction of the renal tubule by myoglobin precipitation in association with Tamm-Horsfall protein under acidic conditions. Urate precipitation may also occur, which together leads to intraluminal casts, increased intratubular pressure, and subsequently decreased glomerular filtration rate (5, 7, 8). A second mechanism occurs via the heme group of myoglobin, which can enhance renal vasoconstriction and ischemia through activation of the cytokine cascade (4, 7). The third proposed mechanism is oxidant injury through heme-induced reactive oxygen species such as...
superoxide anion, hydrogen peroxide, or hydroxyl radicals, provoking direct oxidative damage to the renal tissue (4, 7, 8).

To prevent the complication of ARF in cases of rhabdomyolysis, prophylactic treatment with mannitol, sodium bicarbonate, and fluids is given (4, 5, 7, 8). The decision to treat is based on whether the patient is at risk of developing ARF; knowledge of the presence of urine myoglobin (uMb) may be helpful in making this decision (2–4).

Several reviews and commentaries suggest that measurement of uMb is not helpful and should not be used owing to issues of Mb instability and poor test methodologies (1, 5). Historically, the measurement of uMb was cumbersome and inaccurate. These early methods took advantage of the pseudo-peroxidase activity of the heme moiety in myoglobin, but required complete removal of hemoglobin in the urine, which was difficult to achieve (3, 9, 10). The development of immunometric assays for myoglobin allowed more specific measurement and the potential for better predictive ability, such as in establishing a cutoff for development of ARF (1, 3, 11). Despite this method development, the older, less-specific methods continue to be used.

To assess whether the measurement of uMb, by any method, aids the prediction of ARF in individuals suspected of having rhabdomyolysis, we conducted a systematic review of the literature.

Materials and Methods

SEARCH STRATEGY
We performed a search of the databases Ovid Medline, Ovid Medline in process, AMED, CINAHL, and EMBASE using the search terms myoglobin$ AND (renal OR ARF OR kidney) restricted to publications from January 1980 to December 2006. The results of the search were uploaded into SRS 4.0 (TrialStat Corporation), a Web-based program for conducting systematic reviews.

STUDY SELECTION
We examined studies that related the measurement of uMb to the outcome of ARF in populations with suspected rhabdomyolysis. We included primary studies, published in English, containing 3 or more human subjects of any age, sex, or ethnicity. We excluded studies in which myoglobinuria was a requirement for entry into the study. Two reviewers, using predefined criteria, independently evaluated eligibility of each study. Disagreements were resolved by discussion with 1 or more investigators.

DATA EXTRACTION AND SYNTHESIS
Data extracted included type of study, number of subjects, age, sex, causes of rhabdomyolysis and diagnostic criteria for ARF, methodology for urine myoglobin testing, and measures of kidney function. We applied descriptive statistics to summarize the data. When the studies presented sufficient information, we generated 2-by-2 tables. The data from studies including >20 patients were used to calculate sensitivity, specificity, and 95% CIs.

ASSESSMENT OF METHODOLOGICAL QUALITY
To assess the quality of the primary studies, we used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool (questions 1–13) (12). The quality items were considered individually rather than as a composite score, as recommended by the developers of this tool.

Results

STUDY SELECTION
Fig. 1 provides a flowchart of the review selection process. Our systematic search identified 1602 article abstracts, from which 1361 articles were excluded through title and abstract screening; full text assessment led to further exclusion of 189 articles. Fifty-two studies fulfilled all inclusion criteria for data extraction. Excluding case series, 8 studies contained sufficient information to construct 2-by-2 contingency tables (Table 1).

STUDY CHARACTERISTICS
The study designs included 32 case series (13–44), 16 cohort studies (11, 45–59), and 4 case control studies (60–63), for a total of 2399 subjects. Of the 52 studies, 16 measured uMb in only a portion of their study group. In the group of patients that had uMb measurement (n = 1341), 581 were positive for uMb and 745 developed acute renal failure. In 42 of the 52 studies that reported sex, there were more male (71%) than female patients. Most of the studies (n = 46) were in the adult population; 6 studies focused on the pediatric population (13, 15, 34, 41, 51, 61). ARF attributed to rhabdomyolysis presented in 35% of the adults and 20% of the children.

The majority of the studies used more than 1 criterion for the diagnosis ARF (63%). The most common measure was high serum creatinine (n = 40, 77%), low urine output (n = 24, 46%), or their combination (n = 19, 37%). A small number of studies based the diagnosis on the requirements of dialysis (n = 11, 21%) or reported low glomerular filtration rate (GFR) (n = 7, 13%).
Various methods were used to detect uMb. Seventeen studies did not report the assay method used (15, 16, 20–22, 24, 25, 30, 36, 37, 39–41, 49, 50, 55, 63). Eighteen studies (35%) employed a colorimetric method, which usually involved a preseparation step by centrifugation or filtration to remove hemoglobin, followed by a commercial dipstick for hemoprotein (11, 14, 17, 26, 27, 29, 33–35, 38, 42, 45, 48, 52–54, 56, 57), and 19 (37%) used immunometric methods (11, 13, 18, 19, 23, 28, 31, 32, 38, 43, 44, 46, 47, 51, 58–62). Two studies applied >1 method to detect or quantify uMb (11, 38). Serum creatine kinase (CK), also increased in rhabdomyolysis, was measured in 38 studies (11, 13–28, 30–34, 36–38, 40–45, 47, 52, 56, 57, 61–63), and of these, only 4 had extractable data comparing uMb with CK (45, 47, 51, 59). Serum myoglobin was also measured in addition to urine myoglobin in 11 studies (19, 20, 23, 31, 40, 41, 43, 44, 46, 59, 61), but only 1 study had extractable data for a contingency table (59).

The 52 studies included several causes of rhabdomyolysis. Twenty studies looked at drugs or toxins (therapeutic/drugs of abuse, alcohol, insect venoms, snake venoms, heavy metals) (13, 14, 16, 18, 20–23, 27, 28, 32, 33, 38, 43, 44, 45, 52, 55, 56, 59, 63), 20 referred to trauma (crush injuries, motor vehicle accidents, surgery, physical torture and abuse, long-term confinement without changing position) (11, 15, 17, 23, 26, 27, 29, 35, 38, 40, 41, 43, 44, 47, 55, 56, 62), 12 to physical exertion (strainful muscle exercise, excessive muscle activity, seizures, long-distance running) (11, 23, 25–27, 30, 31, 38, 39, 43, 44, 49), 11 to metabolic disturbance (metabolic syndromes, endocrine disorders, electrolyte imbalances) (11, 15, 19, 23, 27, 35, 41, 43, 44, 45, 55, 56), 9 to burns (thermal injury, high-voltage electrical injury, lightning) (19, 26, 35, 38, 46, 48, 50, 57, 58), 7 to asphyxiation (19, 21, 35, 38, 42, 51, 61), and 11 listed other causes such as hypothermia, hyperthermia, and multiple myeloma (15, 23, 27, 35, 37, 38, 41, 43, 51, 55, 60). The most common causes of rhabdomyolysis were burns due mainly to electrical injuries (31%), drugs or toxins (23%), and trauma (22%).
<table>
<thead>
<tr>
<th>Author, year</th>
<th>Sample</th>
<th>Cause of rhabdomyolysis</th>
<th>uMb cutpoint*</th>
<th>uMb method</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahsan et al. 2001</td>
<td>Random collections; unknown which samples used to establish positive urine myoglobin result</td>
<td>Trauma, MVA, rhabdomyolysis defined by high creatine kinase</td>
<td>Positive</td>
<td>Colorimetric dipstick∗</td>
<td>4</td>
<td>22</td>
<td>0</td>
<td>7</td>
<td>1.0 (0.51–1.00)</td>
<td>0.24 (0.12–0.42)</td>
</tr>
<tr>
<td>Daher et al. 1997</td>
<td>Random urine collections, 3 collections in 1 week; not clear how many per patient were positive</td>
<td>Mild, moderate, and severe tetanus</td>
<td>Unknown</td>
<td>Immunometric (Boehringerwerke)</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>0.47 (0.25–0.70)</td>
<td>0.67 (0.42–0.85)</td>
</tr>
<tr>
<td>Kojima et al. 1985</td>
<td>19- to 24-h urine collection taken on day 3</td>
<td>Birth asphyxia due to RDS, intracranial hemorrhage, seizure, meconium aspiration syndrome, transient tachypnea</td>
<td>&gt;5 µg/L†</td>
<td>Immunometric (RIA)</td>
<td>9</td>
<td>10</td>
<td>0</td>
<td>15</td>
<td>1.0 (0.70–1.00)</td>
<td>0.60 (0.41–0.77)</td>
</tr>
<tr>
<td>Loun et al. 1996</td>
<td>Random urine; highest value taken of all samples collected for each patient (number per patient and time of collection unknown)</td>
<td>Cardiac arrest, dissecting aorta, seizure, skeletal muscle ischemia, sepsis, trauma (gun shot and MVA), infection</td>
<td>&gt;100 µg/L‡</td>
<td>Immunometric (Stratus II)</td>
<td>4</td>
<td>22</td>
<td>0</td>
<td>4</td>
<td>1.0 (0.51–1.00)</td>
<td>0.15 (0.06–0.34)</td>
</tr>
<tr>
<td>Muckart and Abdool-Carrim 1991</td>
<td>Random urine collected at admission</td>
<td>Trauma due to whip-related injuries</td>
<td>Positive</td>
<td>Colorimetric dipstick∗</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>31</td>
<td>0.38 (0.14–0.69)</td>
<td>0.91 (0.77–0.97)</td>
</tr>
<tr>
<td>Muckart et al. 1992</td>
<td>Random urine collected at admission</td>
<td>Trauma due to whip-related injuries</td>
<td>Positive</td>
<td>Colorimetric dipstick∗</td>
<td>4</td>
<td>7</td>
<td>0</td>
<td>53</td>
<td>1.0 (0.51–1.00)</td>
<td>0.88 (0.78–0.94)</td>
</tr>
<tr>
<td>Roberts et al. 1990</td>
<td>24-h urine collection taken on day 1 or 2</td>
<td>Birth asphyxia due to severe shoulder dystocia, intrapartum cord traction with hemorrhage and sepsis</td>
<td>&gt;2 µg/L†</td>
<td>Immunometric (ELISA)</td>
<td>4</td>
<td>7</td>
<td>0</td>
<td>10</td>
<td>1.0 (0.50–1.00)</td>
<td>0.59 (0.51–0.85)</td>
</tr>
<tr>
<td>Wu et al. 1994</td>
<td>Timed urine collections (2- or 8-h); sample with the highest value used</td>
<td>Trauma due to MVA, drug abuse, fracture (femur, pelvis, face), and general skeletal muscle ischemia</td>
<td>&gt;1000 µg/L§</td>
<td>Immunometric (Opus Plus)</td>
<td>6</td>
<td>14</td>
<td>1</td>
<td>20</td>
<td>0.86 (0.49–0.97)</td>
<td>0.59 (0.42–0.74)</td>
</tr>
</tbody>
</table>

* Myoglobin in µg/L may be converted to nmol/L by multiplying by 0.058.
* TP, true positive; TN, true negative; MVA, motor vehicle accident; RDS, respiratory distress syndrome.
† Hb precipitation with ammonium sulfate and supernatant tested for uMb with a dipstick.
‡ Maximum values in 10 healthy newborn infants.
§ No method details provided.
§ Above assay detection limit.
|
Two studies included only infants to find the value for a positive urine myoglobin test. Two chose arbitrary cutpoints; one chose cutpoints close to the limit of detection. Small study sizes. For the 5 quantitative assays, 2 studies reported sensitivities and specificities were wide, reflecting the heterogeneity. For replication. Information on whether unclear or uninterpretable test results were considered was reported in 38% of the studies. The studies used for creating the contingency tables (11, 45, 47, 51, 53, 54, 59, 61), however, satisfied almost all the quality criteria. The 2 studies by Muckart and colleagues (53, 54) did not provide sufficient detail for the reference or index tests, and the study by Ahsan et al. (45) was also incomplete for the reference test. The major quality failure or uncertainty for these studies was disease progression bias, since the times between urine myoglobin assessment and the diagnosis of ARF were unclear (Table 1).

Discussion

The identification of risk factors for ARF in patients with rhabdomyolysis is important to mitigate complications by initiating aggressive preventive strategies, which also carry risk. Factors identified as having potential to differentiate patients at high and low risk for ARF include the amount of muscle damage, base deficit, renal function (GFR, creatinine, urea), and presence of myoglobin in the urine. To obtain a comprehensive overview of literature available on this subject, our search strategy intentionally was not restrictive regarding study design, ARF criteria, etiology, or method of myoglobin measurement. Fifty-two studies fit our search criteria, and of these, 2 studies contained data to create contingency tables. Specifically, these 8 studies prospectively evaluated groups of patients suspected of being at risk for myoglobin-induced ARF.

The 52 studies covered a wide disease spectrum with diverse etiologies, dissimilar diagnostic criteria for ARF and rhabdomyolysis, various methods of uMb measurement, and differences in sample collection and decision cutpoints. Most studies were case series; there were no randomized control trial studies. The QUADAS tool showed that many studies had poor quality with regard to the timing, description, performance, and interpretation of both the index (uMb) and reference tests. The 8 studies used in the calculation of diagnostic performance were of higher quality, having passed almost all the QUADAS criteria, but remained heterogeneous.

The sensitivity of uMb for detecting ARF was 100% in a majority of the 8 selected studies, although specificity was much lower. The value of a screening test is to be able to detect as many true positives as
possible. The uMb test appears to be able to do this, but at a cost of having many false-positive (FP) results. The number of FP results equaled or exceeded the number of true-positive results in all studies except 1 (47). Only 3 studies (47, 53, 59) had false-negative results.

Other test parameters such as the likelihood ratio (LR) and positive predictive value (PPV) showed poor performance. The LR indicates the relative likelihood of the same particular test result having occurred in a diseased individual vs a nondiseased individual. The positive LRs (LR+) were <2.5, except in the 2 studies by Muckart and colleagues (53, 54). These studies were almost identical but differed in 1 inclusion criterion: in the second study (54), the inclusion criterion of venous bicarbonate of <17 mmol/L was added. The result was an increase of the LR+ from 4.3 to 8.6. On the other hand, the PPV for this study was only 36%. No study exceeded a PPV of 58%, indicating that the uMb test has only a fair probability of correctly identifying a patient with a positive test result who has ARF. The prevalence of ARF among these 8 studies varied from a low of 6.3% (54) to a high of 26.5% (51), highlighting again the heterogeneity of these studies and raising caution regarding interpretation of the results.

The variation between studies, or bias, can be due to a variety of factors such as study design, disease spectrum, reference standard, test methodology, and chance (small sample size) (64). The 8 selected studies did not include a large number of patients, with sample sizes ranging from 64 in the largest study (54) to 21 in the smallest (61).

In addition, varying etiologies of disease observed across the studies could also be a bias and explain some of the differences seen in this systematic review. The 2 pediatric studies included infants who suffered from asphyxia and were also quite similar with respect to patient number, type of urine collection (24 h), and method of myoglobin measurement (immunoassay). Although the diagnostic criteria for ARF were more restrictive in the study of Kojima et al. (51) than in Roberts et al. (61), the sensitivity and specificity found for uMb in these studies were equivalent. The etiology of the predominantly adult studies varied widely. There were no studies that included primarily patients with burns or drugs, the most predominant causes of rhabdomyolysis (3, 6, 7).

In the study that examined only patients who had tetanus, the authors demonstrated that adrenergic autonomic overactivity, rather than rhabdomyolysis, was the major contributor to the development of renal failure (47). This might explain why this study had the poorest sensitivity for uMb compared to the other 7 studies.

Higher-risk populations would be expected to have better diagnostic test accuracy, but in a prospective study (not included in Table 1), which selected only patients diagnosed with rhabdomyolysis, the sensitivity was only 78% and specificity 20%. Therefore, other factors were involved, leading to poor test performance (35).

The method used to assay urine myoglobin is another variable in the results of diagnostic studies. Limitations in colorimetric dipstick methods and immunoassays are widespread in the literature (1, 3, 6, 11, 38). However, the detection of Mb in urine depends also on physiological factors, which may go unappreciated and can be difficult to control or determine. Myoglobin appears in the systemic circulation within a few hours after muscle damage and declines rapidly, with a serum half-life of 2–3 h (1, 3, 5, 8). Physiological variables such as patient hydration status, acid–base status, renal function, and hypoxia affect Mb metabolism and consequently its presence in urine (1, 2, 5). These characteristics were poorly reported in the studies we reviewed. In the study by Kojima et al. (51), however, patients with the highest uMb concentration also had acidemia, and 9 of the 11 patients in this group developed ARF. Similarly, the follow-up study by Muckart et al. (54) also had fewer false-negative (FN) cases when patients with acidemia were selected compared to unselected cases in the previous study by Muckart and Abdool-Carrim (53). It has been postulated that volume depletion and acidemia are necessary for myoglobinuric renal failure to occur (5). Many patients in the study by Wu et al. (59) were prophylactically treated with urine alkalinization, mannitol, and volume expansion. This treatment may have contributed to the number of FP cases reported, since early aggressive treatment may prevent the progression from myoglobinuria to ARF.

Furthermore, sample timing, glomerular filtration rate, and urine flow rate are important preanalytical issues. Urine myoglobin concentration may be falsely low if the collection represents a mixture of urine filtered after the injury with preinjury urine from the bladder. Alternatively, the presence of uMb may indicate good renal function in that the kidneys are still able to excrete the high plasma Mb load. As renal function deteriorates, less myoglobin will appear in the urine.

The details on how urine samples were collected was not always clear in the studies we reviewed. The detection of uMb depends on how the sample is collected (random or 24 h) and storage conditions, in addition to the analytical method used for its detection. Myoglobin is unstable in the low pH of the urine (1, 3, 59). Wu et al. (59) demonstrated that Mb in urine samples at pH 5.5 degrades by approximately 35% in 24 h and entirely within a few days, even when stored at −70 °C, whereas de Waard and van ’t Sant
confirmed that at pH 4.5 the Mb concentration in urine was reduced to <10% in <2 h. Only 3 of the 8 studies documented taking precautions to stabilize myoglobin in the urine (11, 45, 59).

This systematic review found that colorimetric dipstick and immunometric assays were used in similar proportions. In contrast with the dipstick methods, immunometric methods have the advantage of allowing a quantitative measure to be made without interference by hemoglobin. Neither of these measurement approaches was developed for the primary use of quantifying myoglobin in urine, however, and both approaches suffer from a number of limitations.

The dipstick method was developed to measure blood in the urine and does not differentiate between hemoglobin, myoglobin, or red blood cells (1, 3). The presence of hemoglobin is common in cases of rhabdomyolysis, and in 1 study hematuria was reported as being present in 32% of the patients (6). The addition of a separation step to remove hemoglobin allows the preferential detection of myoglobin. During the ammonium sulfate step, Mb may coprecipitate with the hemoglobin, resulting in a FN result, or hemoglobin may not precipitate completely and give a FP result (1, 3, 4). Falsely low or FN results can also occur in the presence of ascorbic acid, high nitrite concentration, or high specific gravity, whereas bleach or bacterial peroxidases lead to FP reports (1, 3). Because of the inconsistency in separation (even using ultrafiltration (9, 10) and interferences, this method is not recommended (1, 4).

Immunometric assays were developed for measurement of serum myoglobin for diagnosis of myocardial infarction. In a recent study, 5 commercial immunoassays were compared and found to vary in their ability to detect myoglobin in urine (65). The choice of cutpoint also becomes an issue with these quantitative assays. Two studies that used an immunometric assay chose any detectable amount as being positive, whereas other studies selected higher cutpoints. Loun et al. (11) showed a drop in the number of FP results from 22 to 13 when the cutpoint was increased from 100 µg/L to 20 000 µg/L, but true-positive results remained the same.

The lack of quality evidence to support the use of uMb as a predictor of ARF in patients with suspected rhabdomyolysis raises concerns regarding the value of this test. Since our systematic review was completed, no other studies have been published that would have met our inclusion criteria or added data to our study.

Future studies need to be designed adequately, with larger sample sizes, standardized collection protocols, validated quantitative uMb methods, and comprehensive data collection. ROC analysis will provide information on the optimal uMb cutpoint, and multivariate regression analysis will determine which variables affect the predictive ability of uMb for ARF. Studies should be blinded prospective observational studies. Patients with, or at risk for, rhabdomyolysis that may lead to ARF should have uMb measured at least daily for as long as the risk remains. The outcomes of interest are the diagnosis of ARF or the institution of therapy to prevent ARF within a time period from the uMb result that is clinically appropriate. If these future studies yield promising results, randomized clinical trials might be undertaken where all patients receive the uMb test but are randomized as to whether the results are released.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures of Potential Conflicts of Interest: Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:

Employment or Leadership: None declared.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
 Honoraria: None declared.
Research Funding: This study was supported by a research grant from the Academic Initiatives Program, Department of Pathology and Molecular Medicine, McMaster University to C. Balion.
Expert Testimony: None declared.
Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: The authors thank V. Chetty for early discussions on the design of this systematic review and Andrew Don-Wauchope for his critical reading of the manuscript.

References