Global *emm* type distribution of group A streptococci: systematic review and implications for vaccine development

Andrew C Steer, Irwin Law, Laisiana Matatolu, Bernard W Beall, Jonathan R Carapetis

***emm* sequence typing is the most widely used method for defining group A streptococcal (GAS) strains, and has been applied to isolates in all regions of the world. We did a systematic review of the global distribution of GAS *emm* types. 102 articles and reports were included (38 081 isolates). Epidemiological data from high-income countries were predominant, with sparse data from low-income countries. The epidemiology of GAS disease in Africa and the Pacific region seems to be different from that in other regions, particularly high-income countries. In Africa and the Pacific, there were no dominant *emm* types, a higher diversity of *emm* types, and many of the common *emm* types in other parts of the world were less common (including *emm*1, 4, 6, and 12). Our data have implications for the development of GAS vaccines. On the basis of the available data, the current formulation of the experimental multivalent *emm* vaccine would provide good coverage in high-income countries, particularly USA, Canada, and Europe, but poor coverage in Africa and the Pacific, and only average coverage in Asia and the Middle East.**

Lancet Infect Dis 2009; 9: 611–16

Centre for International Child Health, University of Melbourne, Australia (A C Steer FRACP, I Law MAE); **Fiji Group A Streptococcal Project, Ministry of Health, Suva, Fiji** (L Matatolu BSc); **Streptococcus Laboratory, Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA** (B W Beall PhD); **and Menzies School of Health Research and Charles Darwin University, Darwin, Australia** (J R Carapetis PhD)

Correspondence to: Andrew Steer, Centre for International Child Health, University of Melbourne, Department of Paediatrics, Flemington Road, Parkville, VIC 3052, Australia
andrew.steer@rch.org.au

Introduction

Group A streptococcal (GAS) infections are a major cause of morbidity and mortality worldwide.[1] *Streptococcus pyogenes* causes a wide range of clinical disease. In high-income countries, pharyngitis and invasive disease are the GAS diseases of greatest public health importance, whereas in low-income countries, acute rheumatic fever, rheumatic heart disease, invasive disease, and acute post-streptococcal glomerulonephritis are the major severe diseases, with endemic streptococcal impetigo also leading to very high morbidity. On a global scale, the overwhelming burden of GAS disease is found in low-income countries, where more than 95% of the estimated 663 000 cases of invasive GAS disease and more than 95% of the estimated 294 000 deaths due to rheumatic heart disease occur. However, accurate data are not available from most low-income countries, and published summary data are likely to be underestimates.[1]

Due to the size and severity of the burden of GAS disease, epidemiological surveillance has been crucial to detect changes in disease distribution in various populations. An important part of epidemiological surveillance for GAS disease has been the typing of collected bacterial isolates. Several different methods are available to type GAS.[2] Typing based on the M protein, a cell-surface protein that is the major virulence and immunological determinant of GAS, has been the most widely used method.[3–5] Classic M-protein serological typing was largely replaced by sequence typing of the 5´ end of the M protein (*emm*) gene in the late 1990s.[6] Large epidemiological studies of pharyngitis and invasive disease have been done using *emm* sequence typing, particularly in the USA, Canada, and Europe.[7–10] Population-based studies using *emm* typing have also been done in many other countries.

Available molecular epidemiological data have informed the development of GAS vaccine candidates. Several vaccine candidates have shown promise; however, only one vaccine, a 26-valent M-protein-based vaccine, has recently reached clinical trials.[11,12] Serotypes for this vaccine were chosen if they were known to be common causes of invasive GAS disease or uncomplicated pharyngitis in the USA, or if they were associated with rheumatic fever in classic studies from the USA in the mid-20th century.[13] Recent studies of GAS disease in North America found that *emm* types in the 26-valent vaccine accounted for 79% of all invasive isolates from ten sites in the USA between 2000 and 2004, and 85% of pharyngitis isolates from 13 sites in the USA and Canada between 2000 and 2007.[7,14]

Differences in the distribution of *emm* sequence types between global regions have been noted previously,[7,15] but a thorough review of the available global data has not been undertaken, as it has for other bacteria, including *Streptococcus pneumoniae*.[16,17] In addition, a review of the coverage and potential impact of the experimental multivalent GAS vaccine on a global scale, particularly in low-income countries where the burden of disease is greatest, has not been undertaken. Therefore, we did a global review of the distribution of *emm* types of GAS and assessed the implications of our findings for the development of GAS vaccines.

Methods

Data sources

We searched for studies that described the epidemiology of GAS based on *emm* or M typing by use of a systematic approach that complied with the QUORUM guidelines.[18] Figure 1 summarises our approach. Searches were done in Medline and EmBase from the start of 1990 to the end of March, 2009, by use of the search term "*Streptococcus pyogenes*" combined with the search terms "epidemiology", "*emm*", and "streptococcal M protein". Relevant abstracts from the Lancefield Symposia on Streptococci and Streptococcal Infections held between 1990 and 2008 were also reviewed. No language restrictions were used in the initial search.

Study eligibility and quality assessment

All abstracts from the initial search were checked for relevance (by ACS). Studies that were clearly not population based or studies that did not report *emm* or M typing were excluded. Review articles and studies not dealing with

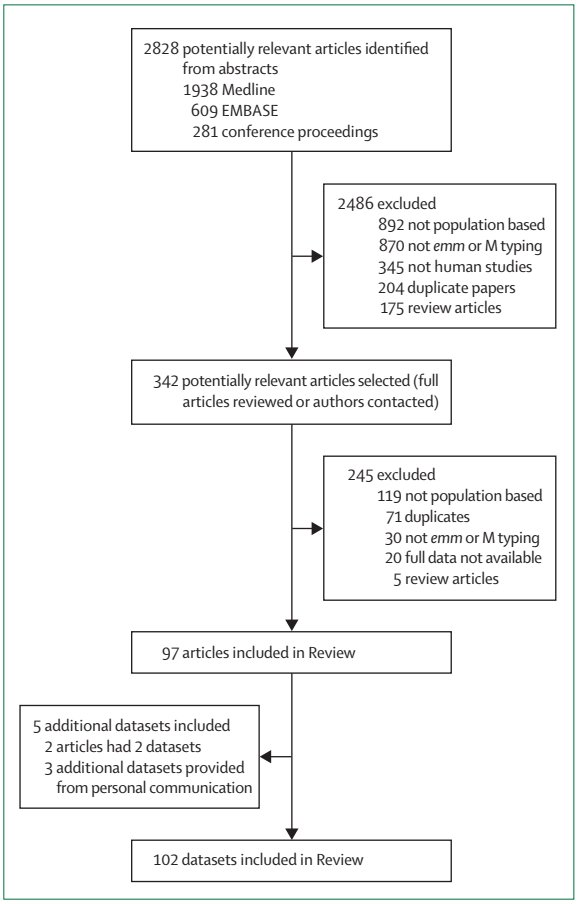

Figure 1: **Summary of selection process and reasons for exclusion of studies**

human beings were excluded. After abstracts had been screened, full papers were retrieved for all remaining articles. Studies could be of retrospective or prospective design, but had to be representative of the population in which they were based. Therefore, articles that dealt with outbreaks of disease and articles that reported *emm* typing of antibiotic-resistant isolates only were excluded. Articles that contained overlapping data already contained in other articles or reports were excluded. In the event of incomplete data in any of the final set of articles, we contacted the investigators directly to obtain a complete dataset if possible; the dataset was only included if it was complete.

Sequence typing

We updated *emm* sequence type information provided in some of the older studies by use of the Centers for Disease Control and Prevention *emm* database, because designation of *emm* sequence types has been continually updated over time (eg, the designation st3365 has been replaced by the designation *emm*119).[19] We used published data to infer *emm* types from M types.[20] Isolates that were not able to be *emm* typed in the referenced articles were included in our study and classified as "nontypable", whereas isolates reported as "other" in referenced articles were removed from all analyses (n=1663).

Classification by region, clinical specimen type, and time period

Studies were categorised by region based on United Nations Populations Prospects classifications,[21] with some modifications. The first modification was that the definition of high-income nations (Europe, North America, Australia, New Zealand, and Japan) was expanded to include Hong Kong.[1] The second was that Indigenous Australians were included in the Pacific region because of documented epidemiological similarities between this group and Pacific Islanders, particularly related to GAS infections.[1]

Isolates were categorised into one of four clinical disease sites/states: invasive, pharyngeal, skin, and other. "Invasive" refers to all isolates that caused invasive disease in the referenced article. "Pharyngeal" refers to all isolates specified as causing pharyngitis or throat carriage, as well as pharyngeal isolates that were not clearly designated. "Skin" refers to all isolates that caused impetigo or other skin infection not otherwise specified in the referenced article. "Other" refers to isolates taken from patients with acute rheumatic fever, acute post-streptococcal glomerulonephritis, scarlet fever, and tic disorders. Isolates of a clinical type not clearly specified in the referenced article were classified as "not differentiated". These "other" and "not differentiated" isolates were included in the overall and regional analyses. For disease-specific analyses, only isolates that were clearly designated as being invasive, pharyngeal, or skin isolates were included.

Studies were also categorised into two time periods (1990–99 and 2000–09) for analysis over time. Studies that spanned both these time periods were not included in this analysis.

Statistical analysis

We compared the distribution of *emm* types between global region by ranking them as a percentage of the total number of isolates for the specified region. We also compared the distribution of *emm* types by clinical specimen site and time period within regions by use of the same technique. Simpson's index of diversity was used to measure the variation of the number of *emm* types within a region.[22] The higher the index, the greater the probability that any two randomly selected isolates from the same population will be of different *emm* types (ie, the greater the diversity of *emm* types). 95% CIs for the index of diversity were calculated as previously described.[23]

Vaccine coverage was defined as the proportion of all isolates in the region or the clinical disease state that were covered by the 26-valent M-protein-based GAS vaccine currently under clinical investigation (emm types 1, 2, 3, 5, 6, 11, 12, 14, 18, 19, 22, 24, 28, 29, 33, 43, 59, 75, 76, 77, 89, 92, 94, 101, 114).[12] 95% CI estimation of vaccine coverage was determined by use of robust SEs that assumed individual observations were independent between studies. Data were

entered into Excel (Microsoft, Redmond, WA, USA) and subsequent analysis was done with Stata version 10.1 (StataCorp, College Station, TX, USA).

Results

The final database contained 102 datasets, contributing 38 081 isolates (webappendix). These isolates were not equally distributed between regions, with high-income countries contributing 32 143 isolates (84·4%), Asia contributing 2248 isolates (5·9%), Pacific Island countries and Indigenous Australians contributing 1383 isolates (3·6%), the Middle East contributing 1219 isolates (3·2%), Latin America contributing 757 isolates (2·0%), and Africa contributing 331 isolates (0·9%). 31 334 isolates were designated as being invasive, pharyngeal, or skin isolates, and of these, 17 173 (54·8%) were invasive isolates, 11 953 (38·2%) were pharyngeal isolates, and 2208 (7·0%) were skin isolates. Of the 11 953 pharyngeal isolates, 9258 (77·5%) were collected from cases of pharyngitis.

A total of 205 *emm* types were recorded, including the category of nontypable isolates. The most common *emm* type was *emm*1, which accounted for 18·3% of all isolates in the study, followed by *emm*12 (11·1%), *emm*28 (8·5%), *emm*3 (6·9%), and *emm*4 (6·9%). However, there were significant differences in *emm* type distribution by region and by clinical disease state.

Distribution by region

There were obvious similarities in *emm* type distribution between the high-income countries, Asia, the middle east, and Latin America, by contrast with the distribution in Africa and the Pacific region. Figure 2 compares the proportions of the 25 most common *emm* types in high-income countries, Africa, and the Pacific (further data are available from the authors).[24] In high-income countries, 25 *emm* types accounted for 90·3% of all isolates, and 146 types contributed the remaining 9·7%. In Africa, 26 *emm* types accounted for 62·5% of all isolates, and 65 contributed the remaining 37·5% of isolates. 26 *emm* types accounted for 61·8% of all isolates from the Pacific region, and 74 types contributed the remaining 38·2% of isolates. *emm*1 and *emm*12 were the two most common *emm* types in high-income countries, Asia, and Latin America, and the second and third most common *emm* types in the Middle East, accounting for between 26·1% and 40·0% of all isolates in these regions. By contrast, *emm*1 was ranked fifth in Africa and 13th in the Pacific region, accounting for 3·6% and 2·0% of isolates, respectively. *emm*12 did not appear in the 25 most common sequence types in the Pacific region. Other than *emm*1 and *emm*12, there were several other common *emm* types found in high-income countries, Asia, the middle east, and Latin America; these included *emm*4 and *emm*6, which were ranked among the eight most common *emm* types in all four regions. By contrast, *emm*4 was ranked 32nd in the Pacific region and was not reported in any of the African studies included in the database, whereas

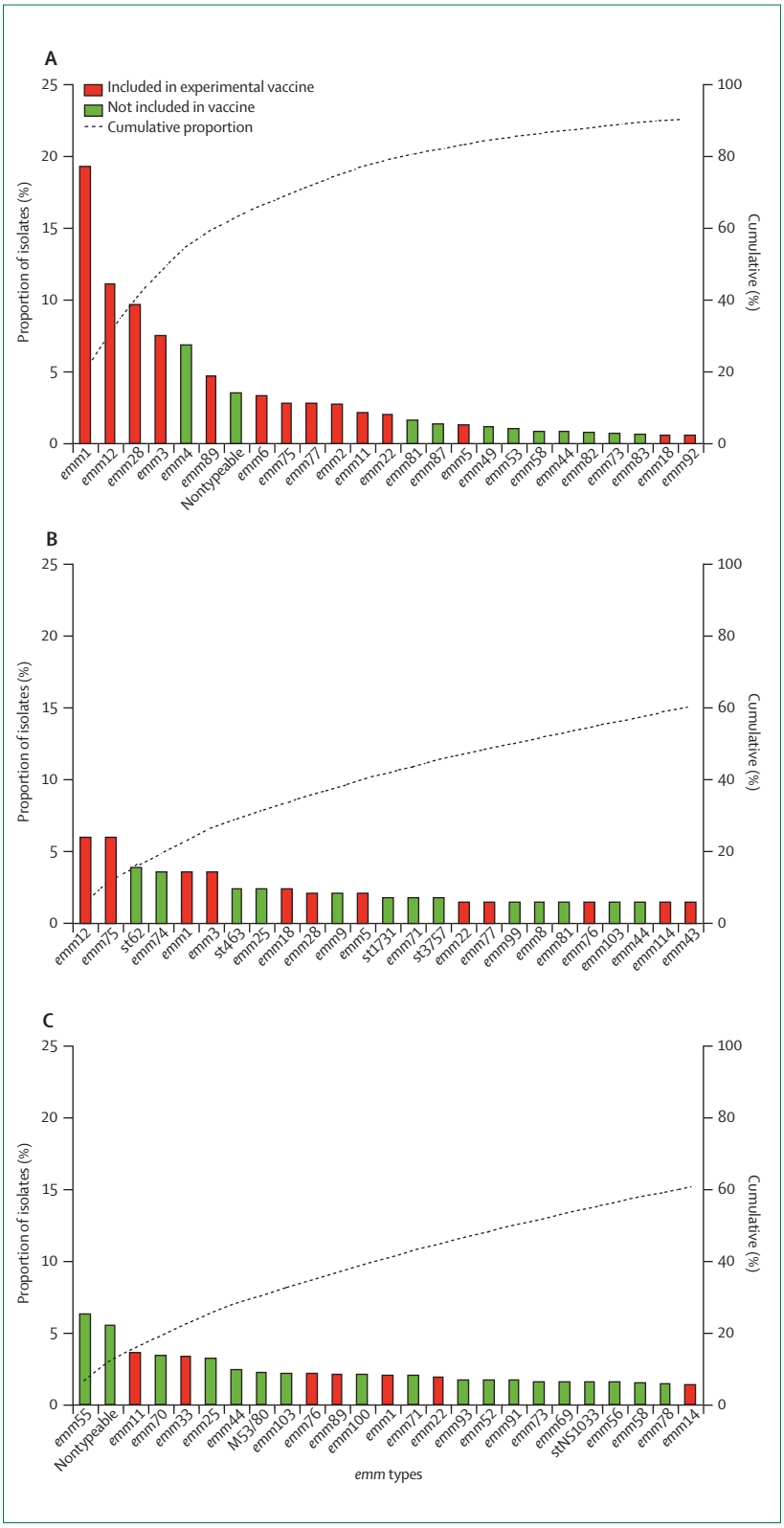

***Figure 2*: 25 most common *emm* types as proportions of all isolates in high-income countries (A), Africa (B), and the Pacific region (C)**
In Africa *emm*112 and in the Pacific region *emm*74 were equal 25th, but are not included.

See Online for webappendix

	Simpson's index of diversity[22] (% [95% CI])
Africa	98·1% (97·7–98·5)
Asia	88·7% (88·0–89·4)
Latin America	93·2% (92·4–94·1)
Middle East	93·0% (92·4–93·5)
Pacific region	97·9% (97·7–98·1)
High-income countries	92·1% (92·0–92·3)
Combined	92·8% (92·7–92·9)

Table 1: **Diversity of *emm* types by global region**

	Invasive	Pharyngeal	Skin	All
Africa	..	43·7% (33·2–54·1)	21·3%*	39·0% (27·2–50·7)
Asia	65·8% (47·4–84·1)	52·5% (39·3–65·7)	41·1% (37·2–44·9)	60·5% (50·8–70·1)
Latin America	77·7% (62·6–92·8)	73·7% (49·6–97·8)	27·0% (19·6–34·4)	71·9% (46·8–97·0)
Middle East	67·5% (57·3–77·7)	54·7% (26·6–82·8)	..	63·2% (50·5–76·0)
Pacific	32·1% (14·4–49·8)	30·8% (26·9–34·6)	19·3% (10·4–28·2)	23·9% (17·8–29·9)
High-income countries	74·9% (71·1–78·7)	77·8% (68·1–87·5)	36·8% (14·5–59·0)	72·8% (66·4–79·2)
Combined	74·2% (70·5–77·9)	73·5% (61·4–85·5)	30·6% (18·3–42·8)	69·7% (63·1–76·2)

Data are vaccine coverage [% (95% CI)]. *95% CI not calculated because data are from one study only. ..=Data not available.

Table 2: **Vaccine coverage of isolates by region and disease**

*emm*6 was ranked 29th in Africa and was not reported in any of the Pacific studies.

Distribution by clinical specimen site

No data were available for invasive disease in Africa nor for skin disease in the Middle East. There was substantial overlap in common *emm* types found in invasive and pharyngeal isolates both within and between the Asian, Latin American, Middle-East, and high-income countries (further data are available from the authors).[24] However, there was less overlap when common *emm* types found in skin isolates were compared with common *emm* types found in pharyngeal and invasive isolates in these regions. For example, in high-income countries, among the ten most common *emm* types found in pharyngeal, invasive, and skin isolates, eight *emm* types were shared between invasive and pharyngeal isolates, whereas five *emm* types were shared between skin and pharyngeal isolates, and five between skin and pharyngeal isolates. In Asia, there were five *emm* types shared between the ten most common *emm* types found in invasive and pharyngeal isolates, but only one type shared between skin and invasive isolates, and three types shared between skin and pharyngeal isolates. In the Pacific region, there were four *emm* types shared among the ten most common *emm* types found in invasive and pharyngeal isolates, two types shared between skin and invasive isolates, and two types shared between skin and pharyngeal isolates. In Africa, two *emm* types were shared among the ten most common types found in skin and pharyngeal isolates.

Distribution over time

The period 1990–99 included 40 unique studies, as did 2000–09. 22 unique studies were excluded from the analysis because they overlapped the time periods. Seven of the ten most common *emm* types found in high-income countries during 1990–99 were also among the ten most common during 2000–09, with remarkably few changes in the order of prevalence. There was less overlap in Asia (four in the ten most common in both periods), Latin America (four), Africa (three), and the Pacific (one). No data were available for the middle east in the period 2000–09.

Diversity

There was greater diversity of *emm* types in Africa and the Pacific than in other regions, as indicated by the smaller proportions of isolates accounting for the first 25 *emm* types in these regions, and by the straighter cumulative curve (figure 2). A few *emm* types accounted for more than 50% of isolates in Asia, Latin America, the Middle East, and high-income countries: in Asia, three *emm* types accounted for 54·0% of isolates; in Latin America, six *emm* types accounted for 54·2% of isolates; in the Middle East, five *emm* types accounted for 53·8% of isolates; and in high-income countries, five *emm* types accounted for 54·5% of the isolates. By contrast, 18 *emm* types accounted for 50·4% of isolates in Africa and 19 *emm* types accounted for 51·3% of isolates in the Pacific region. These differences in diversity were also indicated by calculation of Simpson's index (table 1): the index was higher in Africa and the Pacific compared with the other regions, and this difference was significant on the basis of the 95% CIs.[23]

Vaccine coverage

Table 2 summarises theoretical coverage of isolates by region and disease by the experimental multivalent vaccine.[12] Although there was some heterogeneity between studies within regions, there were clear regional patterns of vaccine coverage. Overall, the 26 *emm* types in the experimental multivalent vaccine accounted for less than 65% of all isolates in four out of the six regions (Africa, Asia, Middle East, Pacific region), and was particularly low in Africa and the Pacific region. Of the disease states, coverage by the vaccine was best for invasive disease and worst for skin disease.

Discussion

Our study has revealed differences in the *emm* type distribution of GAS across global regions, and in particular has revealed marked differences in the molecular epidemiology in Africa and the Pacific region compared with high-income countries. A distinct profile of *emm* types exists in Africa and the Pacific, with an apparent lack of dominant *emm* types and a greater molecular diversity.

The reasons for the contrasting molecular epidemiology in Africa and the Pacific are not clear. It might be that differing clinical presentations of GAS infection in these regions contribute to differing *emm* type profiles. GAS

impetigo is endemic in many parts of Africa and the Pacific, but is much less common in high-income countries.[25–28] The high burden of impetigo in tropical countries is accompanied by large numbers of circulating GAS of multiple *emm* types that are readily transmitted.[25,29,30] Studies of site tropism for GAS in tropical countries have found that most circulating *emm* types of GAS are of *emm* pattern D (skin tropism) or E (both skin and pharyngeal tropism), as opposed to temperate regions where there are more strains of *emm* pattern A–C (pharyngeal tropism).[15] Rapid transmission, negative selection pressure due to a seemingly weak immune response to GAS in the skin,[31] and a lack of effective public-health control measures all support the notion that skin *emm* types dominate the epidemiological scene in many tropical settings.

Our results have implications for vaccine development. The M protein is highly immunopotent, and sera containing serotype-specific antibodies were shown to be protective against re-infection with the same serotype in studies done in the 1950s.[4,32] A 26-valent M protein vaccine has been found to be safe and immunogenic in human beings.[12] The available data suggest that the current formulation would cover most disease-causing serotypes in high-income countries, particularly in USA, Canada, and Europe. However, our results suggest the current formulation of this vaccine would provide limited coverage of disease-causing GAS *emm* types in Africa and the Pacific, and only mid-level coverage in Asia and the Middle East. Alternate formulations with different *emm* type profiles would be needed for different regions, and because of the even distribution of *emm* types in Africa and the Pacific region, a larger number of *emm* types would need to be included in the vaccine in these regions to obtain coverage against a similar proportion of isolates to that in other regions. A further barrier to multivalent M-protein vaccines is the potential for rapid emergence of new *emm* types. Our study found apparently different *emm* profiles over time in most regions, although with a more stable profile in high-income countries. Changes in the predominant circulating *emm* types within relatively short time periods has been shown to occur in individual regions in the USA,[33] although this was explained by the replacement of common *emm* types with other common types from within other regions in the USA, all of which are included in the 26-valent vaccine. Therefore, introduction of a multivalent M-specific GAS vaccine in the USA would seem to be a sound strategy, although close surveillance of prevalent GAS *emm* types within a population would be warranted for an extended period. However, there is potential for rapid and dramatic turnover in regions where there are high numbers of circulating GAS strains (ie, in areas where pyoderma caused by GAS is endemic).[25]

Different approaches to the selection of vaccine antigens, including investigation of vaccines containing conserved epitopes, might be more appropriate in low-income countries, particularly in Africa and the Pacific region, possibly in combination with selected serotype-specific antigens.[31,34–36] Potential conserved antigens are under investigation, including C5a peptidase, GAS carbohydrate, and a peptide within the conserved region of the M protein known as J8.[34–36]

The main limitations to our study were the degree of heterogeneity within some of the regions, the predominance of data from high-income countries, and the potential for selection bias related to time periods in which data were collected. Caution needs to be exercised in interpreting the quantitative comparisons presented in our Review. Although we attempted to choose population-representative studies, there were variations in the methods used by the studies included. Therefore, formal statistical tests of between-study differences were avoided. An indication of study heterogeneity can be found by comparing vaccine coverage proportions and 95% CIs within regions (webappendix). Few data are available from regions that carry the highest burden of GAS disease (ie, Africa and the Pacific).[1] The relatively small number of studies from some regions, particularly Africa, is a potential source of bias and our findings might not be generalisable to all countries in these regions. However, the data reviewed are further evidence of the need for better research into GAS disease in low-income countries, on which we have previously reported.[37,38] Another limitation of our data is that *emm* types are not always adequate strain markers, because they can be shared by unrelated clonal types.[39] Assessment of M-type-specific vaccines against the known array of clones that exist within given *emm* types is thus potentially important, particularly because changes can also occur within a single clone that is represented by a single *emm* type. Such a scenario is underscored by relatively subtle chromosomal changes that have occurred within the highly prevalent *emm*1 clone.[40,41] Finally, a further limitation of our study was that the use of only one reviewer to screen and select the eligible studies might have been a source of selection bias.

In conclusion, our study underscores the need for further molecular epidemiological data from regions where the burden of GAS disease is greatest. However, on the basis of the available data from such regions, the molecular epidemiology of GAS infections in Africa and the Pacific seems to be different from that in other regions, such as high-income countries. This might be related to the high prevalence of GAS impetigo in regions with large numbers of circulating strains. Although the current formulation of the experimental multivalent vaccine provides good coverage of disease-causing *emm* types in most high-income countries, this vaccine would provide poor coverage in Africa and the Pacific, and only average coverage in Asia and the Middle East.

Search strategy and selection criteria

These are described in detail in the Methods section.

Contributors

ACS was the primary coordinator of data collection, analysis, and writing. IL was primarily involved in data analysis and writing. LM was primarily involved in data collection and writing. BB and JRC supervised data collection, analysis, and writing. All authors contributed substantially to the preparation of the paper.

Conflicts of interest

There are no relevant conflicts of interest to disclose for any author.

Acknowledgments

The following investigators provided additional *emm* typing data: Robert Tanz, Northwestern University, Chicago, IL, USA; Chris Van Beneden, Centers for Disease Control and Prevention, Atlanta, GA, USA; Marc Levy, Centre Hospitalier de Polynésie Française, Papeete, Tahiti; Rajesh Kumar, Post Graduate Institute of Medical Education and Research, Chandigarh, India; Tuula Siljander, National Public Health Institute, Mannerheimintie, Finland; Tadayoshi Ikebe, National Institute of Infectious Diseases, Tokyo, Japan; Guliz Erdem, John A Burns School of Medicine, University of Hawaii, Manoa, Hawaii, USA; Graham Magor, Queensland Institute of Medical Research, Brisbane, Australia; and Kerry-Ann O'Grady and Leisha Richardson, Menzies School of Health Research, Darwin, NT, Australia.

References

1 Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. *Lancet Infect Dis* 2005; **5:** 685–94.

2 Johnson DR, Kaplan EL, Sramek J, Bicova R, Havlicek J, Havlickova H. Laboratory diagnosis of group A streptococcal infections. Geneva: WHO, 1996.

3 Lancefield RC. The antigenic complex of *Streptococcus hemolyticus*, I: demonstration of a type-specific substance in extracts of *Streptococcus hemolyticus*. *J Exp Med* 1928; **47:** 9–10.

4 Lancefield RC. Current knowledge of the type specific M antigens of group A streptococci. *J Immunol* 1962; **89:** 307–13.

5 Beall B, Facklam R, Thompson T. Sequencing *emm*-specific PCR products for routine and accurate typing of group A streptococci. *J Clin Microbiol* 1996; **34:** 953–58.

6 Facklam R, Beall B, Efstratiou A, et al. *emm* typing and validation of provisional M types for group A streptococci. *Emerg Infect Dis* 1999; **5:** 247–53.

7 O'Loughlin RE, Roberson A, Cieslak PR, et al. The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000–2004. *Clin Infect Dis* 2007; **45:** 853–62.

8 Lamagni TL, Efstratiou A, Vuopio-Varkila J, Jasir A, Schalen C. The epidemiology of severe *Streptococcus pyogenes* associated disease in Europe. *Eurosurveillance* 2005; **10:** 179–84.

9 Shulman ST, Tanz RR, Kabat W, et al. Group A streptococcal pharyngitis serotype surveillance in North America, 2000–2002. *Clin Infect Dis* 2004; **39:** 325–32.

10 Tanz RR, Shulman ST, Kabat W, et al. Five-year group A streptococcal pharyngitis serotype surveillance in North America, 2000–2005. Proceedings of the XVIth Lancefield International Symposium on Streptococci and Streptococcal Diseases, Palm Cove, Australia; Sept 25–29, 2005: 30–33.

11 Kotloff KL, Dale JB. Progress in group A streptococcal vaccine development. *Pediatr Infect Dis J* 2004; **23:** 765–66.

12 McNeil SA, Halperin SA, Langley JM, et al. Safety and immunogenicity of 26-valent group A streptococcus vaccine in healthy adult volunteers. *Clin Infect Dis* 2005; **41:** 1114–22.

13 Hu MC, Walls MA, Stroop SD, Reddish MA, Beall B, Dale JB. Immunogenicity of a 26-valent group A streptococcal vaccine. *Infect Immun* 2002; **70:** 2171–77.

14 Shulman ST, Tanz RR, Kabat W, et al. Seven year surveillance of streptococcal pharyngitis *emm* types in North America. Proceedings of the XVII Lancefield Symposium on Streptococci and Streptococcal Diseases, Porto Heli, Greece; June 22–26, 2008. Abstract O1.3.

15 Bessen DE, Carapetis JR, Beall B, et al. Contrasting molecular epidemiology of group A streptococci causing tropical and nontropical infections of the skin and throat. *J Infect Dis* 2000; **182:** 1109–16.

16 Hausdorff WP, Bryant J, Kloek C, Paradiso PR, Siber GR. The contribution of specific pneumococcal serogroups to different disease manifestations: implications for conjugate vaccine formulation and use, part II. *Clin Infect Dis* 2000; **30:** 122–40.

17 Hausdorff WP, Bryant J, Paradiso PR, Siber GR. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis* 2000; **30:** 100–21.

18 Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999; **354:** 1896–900.

19 Centers for Disease Control and Prevention. *Streptococcus* Laboratory. Protocol for *emm* typing. http://www.cdc.gov/ncidod/biotech/strep/protocol_emm-type.htm (accessed Aug 11, 2009).

20 Facklam RF, Martin DR, Lovgren M, et al. Extension of the Lancefield classification for group A streptococci by addition of 22 new M protein gene sequence types from clinical isolates: *emm*103 to *emm*124. *Clin Infect Dis* 2002; **34:** 28–38.

21 United Nations Population Division. World Population Prospects: The 2008 revision population database. Definition of major areas and regions. http://esa.un.org/unpp/index.asp?panel=5 (accessed Aug 12, 2009).

22 Simpson E. Measurement of diversity. *Nature* 1949; **163:** 688.

23 Grundmann H, Hori S, Tanner G. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *J Clin Microbiol* 2001; **39:** 4190–92.

24 Centers for Disease Control and Prevention *Streptococcus* Laboratory. *emm* types as proportions of total disease isolates in six global regions. http://www.cdc.gov/ncidod/biotech/strep/emmtype_proportions.htm (accessed Aug 11, 2009).

25 Carapetis JR, Currie BJ, Hibble M, Sriprakash KS, Bessen DE, Mathews JD. Rapid turnover of multiple strains of group A streptococcus in an Australian Aboriginal community. In: Martin DR, Tagg JR, eds. Streptococci and streptococcal diseases: entering the new millennium. Auckland: SecuraCopy, 2000: 155–58.

26 Carapetis JR, Currie BJ, Matthews JD. Multiple strains of *Streptococcus pyogenes* in skin sores of Australian Aborigines. *J Clin Microbiol* 1995; **33:** 1471–72.

27 Kristensen JK. Scabies and pyoderma in Lilongwe, Malawi. Prevalence and seasonal fluctuation. *Int J Dermatol* 1991; **30:** 699–702.

28 Masawe AE, Nsanzumuhire H, Mhalu F. Bacterial skin infections in preschool and school children in coastal Tanzania. *Arch Dermatol* 1975; **111:** 1312–16.

29 Ferrieri P, Dajani AS, Wannamaker LW, Chapman SS. Natural history of impetigo. I. Site sequence of acquisition and familial patterns of spread of cutaneous streptococci. *J Clin Invest* 1972; **51:** 2851–62.

30 Reinstein CR. Epidemic nephritis at Red Lake, Minnesota. *J Pediatr* 1955; **47:** 25–34.

31 Bessen DE, McGregor KF, Whatmore AM. Relationships between *emm* and multilocus sequence types within a global collection of *Streptococcus pyogenes*. *BMC Microbiol* 2008; **8:** 59–71.

32 Lancefield RC. Persistence of type-specific antibodies in man following infection with group A streptococci. *J Exp Med* 1959; **110:** 271–92.

33 Shulman ST, Stollerman G, Beall B, Dale JB, Tanz RR. Temporal changes in streptococcal M protein types and the near-disappearance of acute rheumatic fever in the United States. *Clin Infect Dis* 2006; **42:** 441–47.

34 Sabharwal H, Michon F, Nelson D, et al. Group A streptococcus (GAS) carbohydrate as an immunogen for protection against GAS infection. *J Infect Dis* 2006; **193:** 129–35.

35 Shet A, Kaplan EL, Johnson DR, Cleary PP. Immune response to group A streptococcal C5a peptidase in children: implications for vaccine development. *J Infect Dis* 2003; **188:** 809–17.

36 Batzloff MR, Hayman WA, Davies MR, Zeng M, Pruksakorn S, Brandt ER. Protection against group A streptococcus by immunization with J8-diptheria toxoid: contribution of J8-and diptheria toxoid-specific antibodies to protection. *J Infect Dis* 2003; **187:** 1598–608.

37 Carapetis JR. Rheumatic heart disease in developing countries. *N Engl J Med* 2007; **357:** 439–41.

38 Tibazarwa KB, Volmink JA, Bongani MM. The incidence of acute rheumatic fever in the world: a systematic review of population-based studies. *Heart* 2008; **94:** 1534–40.

39 Beall B, Gherardi G, Lovgren M, Forwick B, Facklam R, Tyrrell G. *emm* and *sof* gene sequence variation in relation to serological typing of opacity factor positive group A streptococci. *Microbiology* 2000; **146:** 1195–209.

40 Cleary PP, LaPenta D, Vessela R, Lam H, Cue D. The present day globally disseminated M1 subclone of group A streptococci differs from other subclones by 70 kilobases of prophage DNA and capacity for high-frequency intracellular invasion. *Infect Immun* 1998; **66:** 5592–97.

41 Sumby P, Porcella SF, Madrigal AG, et al. Evolutionary origin and emergence of a highly successful clone of serotype M1 group A Streptococcus involved multiple horizontal gene transfer events. *J Infect Dis* 2005; **192:** 771–82.