International Health 2021; **13**: 606–614
https://doi.org/10.1093/inthealth/ihaa006 Advance Access publication 16 March 2020

ORIGINAL ARTICLE

Reliability of the Service Trip Audit Tool to assess the quality of short-term medical missions

Christopher Dainton[a,*,1], Christina Gorman[b], William Cherniak[c,d], Lorena Lopez[e] and Charlene H. Chu[b,f]

[a]*McMaster University, 1280 Main St West, Hamilton, ON, Canada, L8S 4L8;* [b]*Department of Public Health Sciences, University of Toronto, 155 College Street Toronto, ON M5T 3M7 Canada;* [c]*Department of Family and Community Medicine, Division of Emergency Medicine, Markham Stouffville Hospital, 381 Church St, Markham, ON L3P 7P3, Canada;* [d]*Bridge to Health Medical and Dental, 491 Lawrence Avenue West Suite 301 M5M 1C7, Toronto, ON;* [e]*Universidad Privada Antenor Orrego, Avenida América Sur 3145, Trujillo 13008, Peru;* [f]*Lawrence S. Bloomberg Faculty of Nursing, 155 College St, Toronto, ON M5T 1P8, Canada*

*Corresponding author: Tel: +1-416-884-7231; E-mail: christopher.dainton@gmail.com

[1]Present address: 1002-8 The Esplanade, Toronto, ON, Canada M5E 0A6.

Received 30 September 2019; revised 27 December 2019; editorial decision 21 January 2020; accepted 22 January 2020

Background: We aimed to assess the adherence of short-term medical missions (STMMs) operating in Latin America and the Caribbean (LAC) to key best practices using the Service Trip Audit Tool (STAT) and to calculate the inter-rater reliability of the data points. This tool was based on a previously published inventory of 18 STMM best practices.

Methods: Programme administrators and recent volunteers from 335 North American organizations offering STMMs in LAC were invited to complete the STAT anonymously online. Adherence to each of 18 best practices was reported as either 'yes', 'no' or 'not sure'. Fleiss' κ was used to assess inter-rater agreement of the responses.

Results: A total of 194 individuals from 102 organizations completed the STAT (response rate 30.4%; 102/335 organizations) between 12 July and 7 August 2017. Reported adherence was >80% for 9 of 18 best practices. For 37 non-governmental organizations (NGOs) with multiple raters, inter-rater agreement was moderate to substantial (κ>0.4) for 12 of 18 best practices.

Conclusions: This is the first study to evaluate adherence to STMM best practices. Such an objective evaluation will be valuable to governments, volunteers and NGO donors who have an interest in identifying high-quality partners. Assessment and monitoring of STMMs through self-audit may be foundational steps towards quality improvement.

Keywords: global health, medical missions, primary care.

Introduction

Evidence-based literature on impact and quality has struggled to keep pace with the ongoing popularity of short-term medical missions (STMMs), with one review[1] finding quantitative methods in only 5% of 1100 publications on STMMs over the last 20 y. Based on one nationwide survey of American physicians, an estimated 16.5% of respondents had volunteered on STMMs in 2012.[2] As such, the conduct of sending organizations has become increasingly concerning to global health advocates,[3,4] who cite the real and potential harms associated with poorly conceived international aid projects. Accordingly, the trend towards increased surveillance of the activities of non-governmental organizations (NGOs) in developing countries led the World Health Organization (WHO) to recently propose minimum standards for foreign relief teams in the context of disasters.[5]

The critical ethical issues surrounding STMMs have been explored extensively in previous literature.[3,6–10] These include concerns with the lack of sustainability and partnerships, inadequate cultural preparation and pre-departure training, poor adherence to clinical best practices and providers practising outside of their scope. Such deficiencies can expose patients to inadequate care, negatively impact the healthcare system in fragile economies and contribute to a culture of dependency. In parallel, the number of STMMs is vast and growing,[11] consumes substantial financial resources[8,11] and involves an organizational landscape that is constantly in flux. Subsequently, well-meaning healthcare professionals may end up spending thousands of dollars for an

experience that they cannot properly vet and has an unclear impact on the recipients of their efforts. A tool for organizations to objectively evaluate the quality of their work would serve as a useful starting point for volunteers looking for a well-planned experience that they can support in good conscience.

Given the current prevalence of STMM volunteering, common and objective standards are necessary to evaluate and compare the quality of the services delivered to patients and host communities. A recent systematic review of 92 descriptive and theoretical papers specifically describes best practice recommendations for STMMs, organizing them using the WHO Health Systems Framework.[6] A more recent review of best practice guidelines found 27 different guidelines in the grey literature,[12] although it remains uncertain how to adequately disseminate these emerging perspectives to prospective volunteers. Rather than standardized reporting measures, current practice for medical professionals and trainees selecting a global health experience with an NGO often involves anecdotal recommendations from veteran to prospective volunteers.[12]

While a recognized objective mechanism for the monitoring and evaluation of STMMs does not currently exist, there has been emerging academic consensus on best practices and a push for top-down legislation to enforce them.[6,7,12,13] Such a proposition is challenging given the limited resources in host countries, limited jurisdiction in sending countries and the often limited political will for such an enterprise. Moreover, the basis for such enforcement is challenged by the dearth of quantitative STMM literature describing current outcomes; the heterogeneous structures of the various religious, secular, educational and for-profit NGOs in low- and middle-income countries; and disagreement among guidelines on the relative importance of various aspects of an ideal STMM.[7] As such, the translation of these policy recommendations into real improvements in the quality of STMMs remains a persistent challenge.

An alternative approach to local enforcement involves bottom-up, self-auditing of STMM projects abroad. One seminal study[8] describes a self-assessment tool designed to measure and promote improvements in quality of care on STMMs, but there has been little to no adoption of this framework in actual practice, despite its growing recognition in the academic literature.[1,7,12] The original six major quality domains (preparedness, sustainability, cost-effectiveness, efficiency, impact and education) provide a context-specific frame for comparing subsequent attempts to address quality of care on STMMs, but the authors acknowledge the inherent limitations of this type of subjective self-assessment.[8]

Given these challenges, and with the goal of maximizing utility for prospective volunteers, an ideal assessment tool must be simple to use, easily understood, objective and reproducible.[14] Furthermore, the acceptability of any quality assessment tool depends not only on its foundation in the existing literature, but also on the integration of relevant stakeholder perspectives.[15] This has been a recognized limitation of previously created tools.[3,4,16] In 2017, Dainton et al.[13] sought to integrate the views of clinicians, academics, NGO administrators and student volunteers through an eDelphi process designed to approve a framework of universally acceptable best practices for STMMs. The current study builds on this framework by adapting its elements to create the Service Trip Audit Tool (STAT), which consists of 18 binary questions intended for STMM self-auditing.

The aims of this pilot study were to describe current STMM practices in Latin America and the Caribbean (LAC) based on responses to the STAT survey and assess the objectivity of the elements of the STAT by calculating the inter-rater reliability of respondents. Compiling a quantitative data set describing current STMM practices in terms of these elements will facilitate more specific hypotheses regarding the quality of STMMs as well as outcome evaluations of interventions, partnerships, policies and quality assessment projects.

Methods

A multiphase methodology was used to pilot and determine inter-rater reliability for the previously developed quality assessment tool. Ethics approval was sought from and waived by Markham Stouffville Hospital.

Development of the STAT

The STAT is an audit tool consisting of 18 STMM best practice elements (Table 1) that was previously content validated by an eDelphi panel of international stakeholders.[13] The candidates for inclusion were selected using a theory synthesis methodology that relied on key literature sources and a recent systematic review of STMM best practices.[6] Each of the 18 elements that reached consensus was rephrased into a forced-choice, theoretically falsifiable binary item (yes, no, not sure) that was then edited for clarity and structure to encourage objectivity of the tool, avoid decision fatigue of respondents and allow calculation of a frequency distribution of the items. Respondents were also given an open-ended opportunity for comments on each question, as well as global comments on their STMM at the end of the survey.

The goal was to enable a previous volunteer or STMM programme administrator to anonymously complete a STAT survey in <5 min. For convenience, the items were classified into six major domains (sustainability, education, efficiency, impact and safety, preparedness and cost-effectiveness) that were based on a previous seminal article,[8] to allow respondents to cognitively organize the items.

Sampling and data collection

The sample consisted of NGOs listed in the medicalservicetrip.com database, which is the largest online database of primary care STMMs operating in LAC and which, at the time of the study, consisted of 335 organizations currently operating primary care STMMs (<1 month duration) in LAC. We extracted the following information from each organizational website: the location of their headquarters in North America, the location(s) served in LAC, the frequency of STMMs to LAC, the setting(s) of their mobile clinics, the number and types of providers and whether the organization was faith-based.

A SurveyMonkey online survey was created using the elements of the STAT, and a hyperlink to the survey was incorporated into the individual organization descriptions found on medicalservicetrip.com. Anonymous respondents were asked to categorize themselves as either programme administrators,

Table 1. Overview of the STAT, an audit tool constructed based on a literature-based, eDelphi, stakeholder-validated framework for best practices on STMMs

Domain	Best practice
Sustainability	The organization has a formal partnership with local health services in the host community.
	The organization has a clear referral process for patients who need higher levels of care.
	In addition to the visiting volunteers, the organization ensures that there is always a local clinician involved in clinical care.
	The organization has a permanent staff member or partner organization in the host community.
Education	The organization builds capacity by helping train host providers, local health workers or community health workers.
	The organization engages in public health work or health promotion in the community.
Efficiency	The organization promotes the visiting clinics to locals by word of mouth or advertisement or uses a clinic location that is already well known to locals.
	The organization has a formal staffing plan describing future needs and a recruitment strategy.
	The organization has a formal triage, priority, appointment or ticketing system in place for patients visiting the clinic.
Impact and safety	The organization solicits written feedback/debriefing from volunteers after the trip is over.
	The organization keeps medical records that are easily accessible to future clinicians.
	The organization provides evidence-based clinical guidelines to volunteers, describing an approach to common diseases in the host community.
Preparedness	Volunteers are pre-screened before being accepted by the organization.
	The organization provides pre-departure training for volunteers (i.e. in-person or online).
	Urine dipsticks, pregnancy tests and glucometers are all available and there is a clear pathway for volunteers to obtain more advanced tests.
	The organization provides written clinical protocols to volunteers (i.e. limiting their practice scope to the care they are licensed to provide at home).
Cost-effectiveness	The financial reports for this organization are transparent and easily available (i.e. via website, annual report, etc.).
	The organization considers and describes any host community costs that are associated with hosting volunteers (i.e. on their website).

medical professional volunteers (including medical trainees) or non-medical volunteers. Respondents were requested to have travelled or worked with the organization within the last year, although the anonymity of the survey precluded confirmation of this.

Each NGO was contacted by a research assistant with a standard e-mail addressed to the programme administrator between 1 June and 1 December 2017, using publicly accessible contact information available on their organizational websites, and was provided with a link to their organizational profile on medicalservicetrip.com and a request to complete a self-audit. The e-mailed introduction included the study purpose, its potential benefit to prospective volunteers and the literature and contact information for the authors. The administrator was invited to forward the link to previous volunteers and other programme administrators, with the objective of obtaining multiple ratings for the same organization. The self-audit was also publicly available to website users. No compensation was offered for completion of the survey.

Three follow-up requests were sent to non-responders according to the Dillman approach to internet surveys, which maximizes response rates through personalized, repeated contact.[17] The final follow-up was sent through the organizational Facebook account, if this was available. One additional follow-up was attempted for survey respondents and requested a second rater for their organization.

Sample size calculation

A sample size calculation was completed using Stata 15 (StataCorp, College Station, TX, USA). Assuming a true reliability of 0.8, the sample size required to detect an excellent reliability of at least R−0.1 with 95% confidence[18] would be 70 NGOs, with at least two raters (κ=2) completing the survey for each NGO. We planned to contact all 340 NGOs contained in the database to complete the online survey and expected a response rate of 25–40% (n=85–136) based on previous similar studies using the same database[19] and typical response rates for external online surveys.[20,21]

Statistical analysis

Statistics were calculated using an online statistical tool.[22] Characteristics of the organizations with one or more responses were compared with those with no responses using Fisher's exact test ($p<0.05$) to determine predictors of NGO engagement with the quality assessment study. Adherence to each of the 18 best practices was reported as 'yes', 'no' or 'not sure', and conflicting data

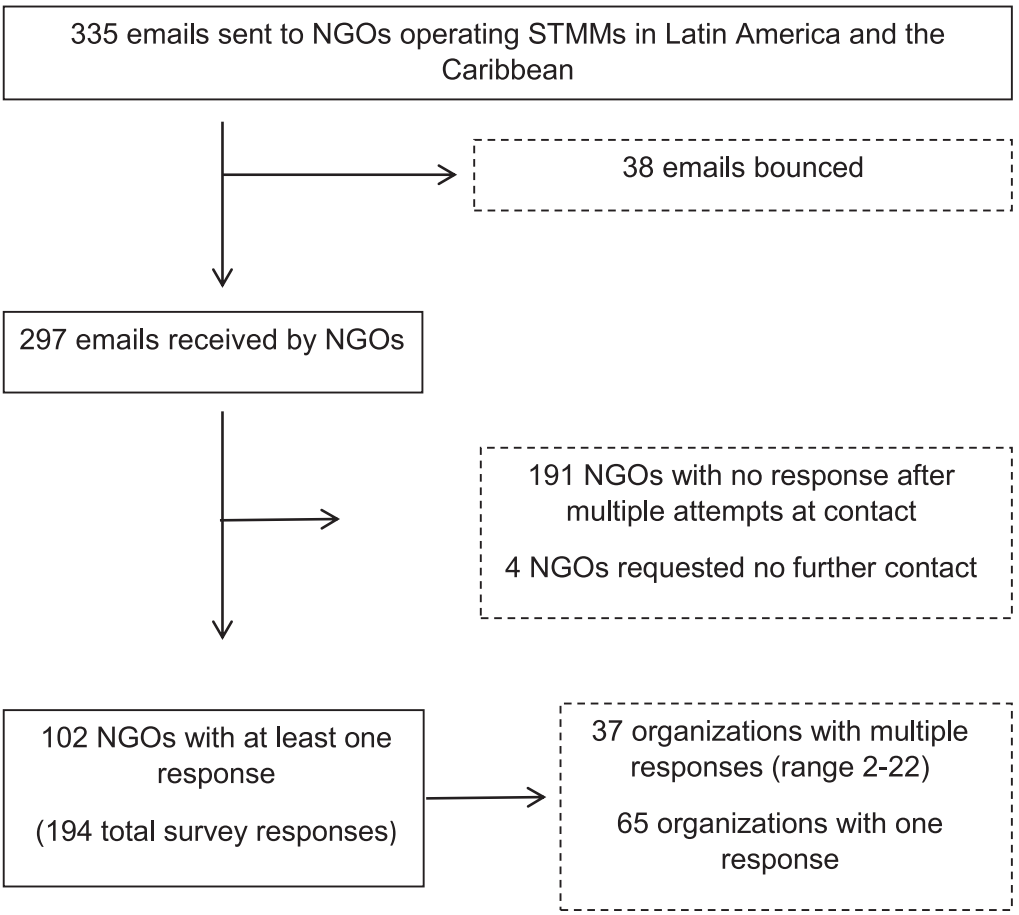

Figure 1. Flow diagram for NGOs operating in LAC who were solicited to complete the STAT for STMMs.

were resolved by investigator consensus. Adherence was calculated as the percentage of organizations with a 'yes' response, with the result being a percentage of organizations claiming compliance with each element of the STAT.

The inter-rater reliability for each STAT question was calculated as a free marginal Fleiss' κ. This measure specifically allows that different organizations may be rated by different individuals in which raters are not forced to assign a certain number of cases to each category.[23,24] Benchmarks of $\kappa > 0.6$ for substantial correlation, $\kappa > 0.4$–0.6 for moderate to good correlation, $\kappa > 0.2$–0.4 for fair correlation and ≤ 0.2 for slight correlation were used.[25] Global comments and comments specific to each element were analysed qualitatively using thematic analysis but are not presented in this study.

Results

Figure 1 describes the flow of e-mail contact with the organizations identified by the study. A total of 335 e-mails were sent out in three rounds beginning 12 July 2017, 7 August 2017 and 21 August 2017. Thirty-eight e-mails bounced back (302 were received). Four organizations requested no further contact.

By 1 March 2018, 194 responses from 97 programme administrators (50%), 57 medical professionals (29.4%) and 40 non-medical volunteers (20.6%) had been received. Responses were submitted from 102 organizations, for an organizational response rate of 30.4%. The characteristics of the 102 responding organizations are described in Table 2 compared with non-responders. Non-responders operated fewer trips per year on average and were more likely to be faith-based, but were similar with regards to type of clinic, length of trip and trip setting.

The claimed adherence of responding organizations to the 18 best practices described by the STAT tool was high (Table 3), with at least 80% of the organizations claiming to comply with 9 of the items in the inventory. The lowest performing inventory items were the presence of minimum diagnostic tests (61.8% of organizations claiming adherence), formal referral processes (65.7%), clinical scope of practice protocols (65.7%), clinical guidelines (58.8%), accessible medical records (65.7%) and community cost–benefit analysis (52.0%).

There were 37 organizations with multiple raters that were included in the reliability analysis. Table 4 indicates the Cohen's κ for each of the 18 items in the STAT inventory. For the organizations with multiple raters, inter-rater agreement was substantial ($\kappa > 0.6$) for 6/18 best practices, moderate ($\kappa > 0.4$–0.6) for 6/18 best practices, fair ($\kappa > 0.2$–0.4) for 4/18 best practices and slight for 2/18 best practices (<0.20).

Table 2. Participants and characteristics of 102 STMMs responding to the STAT versus non-responders

Characteristics		Respondents (N=102), n (%)	Non-respondents (N=233), n (%)	p-Values
Type of organization	Secular	52 (51.0)	90 (38.6)	0.028
	Faith-based	41 (40.2)	133 (57.1)	
	Educational	6 (5.9)	9 (3.8)	
	Unclear	3 (2.9)	1 (1.0)	
Type of clinic[a]	Mobile	72 (70.6)	135 (57.5)	0.968
	Standing clinic	50 (49.0)	100 (39.5)	
	Hospital	19 (18.6)	35 (15.0)	
	Unclear	7 (6.9)	31 (13.3)	
Minimum trip duration	<2 weeks	67 (65.7)	154 (66.1)	0.726
	2–4 weeks	19 (18.6)	34 (14.6)	
	>4 weeks	8 (7.8)	19 (8.2)	
	Unclear	8 (7.8)	26 (11.2)	
Average number of trips per year		10.9	6.1	
Trip setting[a]	Rural	92 (90.2)	186 (79.8)	0.106
	Urban	30 (29.4)	39 (16.7)	
	Unclear	9 (8.8)	37 (15.9)	

p-Values are based on Fisher's exact test, in which unknown data are ignored.
[a]Percentages add to >100% because some organizations operated more than one type of clinic.

Table 3. Reported adherence to 18 best practice elements by 102 STMMs in LAC with at least one completed STAT survey

Best practice element		Yes, n/N (%)	No, n/N (%)	Not sure, n/N (%)
Sustainability	Formal partnership	91/102 (89.2)	10/102 (9.8)	1/102 (1.0)
	Local clinician involvement	82/102 (80.4)	20/102 (19.6)	0/102 (0)
	Permanent staff	87/102 (85.3)	15/102 (14.7)	0/102 (0)
	Formal referral process	67/102 (65.7)	28/102 (27.5)	7/102 (6.8)
Education	Capacity building	80/102 (78.4)	18/102 (17.6)	4/102 (3.9)
	Public health work	88/102 (86.3)	11/102 (10.8)	3/102 (2.9)
Preparedness	Volunteer screening	84/102 (82.4)	15/102 (14.7)	3/102 (2.9)
	Diagnostic tests	63/102 (61.8)	29/102 (28.4)	12/102 (11.7)
	Pre-departure training	83/102 (81.4)	13/102 (12.7)	6/102 (5.9)
	Clinical scope of practice protocols	67/102 (65.7)	27/102 (26.5)	8/102 (7.8)
Efficiency	Promotion of clinics	97/102 (95.1)	3/102 (2.9)	2/102 (2.0)
	Formal triage/scheduling	87/102 (85.3)	12/102 (11.8)	3/102 (2.9)
	Staffing plan	74/102 (72.5)	19/102 (18.6)	9/102 (8.8)
Impact and safety	Written feedback/debriefing	87/102 (85.3)	13/102 (12.7)	2/102 (2.0)
	Clinical guidelines	60/102 (58.8)	33/102 (32.3)	9/102 (8.8)
	Accessible medical records	67/102 (65.7)	27/102 (26.5)	8/102 (7.8)
Cost-effectiveness	Financial transparency	76/102 (74.5)	18/102 (17.6)	8/102 (7.8)
	Community cost–benefit analysis	53/102 (52.0)	36/102 (35.3)	13/102 (12.7)

'Yes' responses for organizations with multiple raters represent majority responses and investigator consensus based on qualitative analysis of comments (not presented in this study).

Discussion

In this pilot study, NGO representatives and volunteers claimed adherence to the majority of the best practices included in the STAT tool, with moderate to substantial interrater reliability for most data points. To our knowledge, this is the first description of a bottom-up, objective approach to quality assessment of STMMs. It represents a novel approach to knowledge translation by encouraging discussion between programme administrators, volunteers and local partners.

Table 4. Inter-rater correlation for each of 18 elements assessed by the STAT for STMMs with two or more responses (n=37), determined by free marginal Fleiss' κ

Element assessed		Fleiss' κ	Inter-rater agreement
Sustainability	Formal partnership	0.779	Substantial
	Local clinician involvement	0.378	Fair
	Permanent staff	0.632	Substantial
	Formal referral process	0.484	Moderate
Education	Capacity building	0.549	Moderate
	Public health work	0.611	Substantial
Preparedness	Volunteer screening	0.634	Substantial
	Diagnostic tests	0.298	Fair
	Pre-departure training	0.583	Moderate
	Clinical scope of practice protocols	0.283	Fair
Efficiency	Promotion of clinics	0.793	Substantial
	Formal triage/scheduling	0.495	Moderate
	Staffing plan	0.278	Fair
Impact and safety	Written feedback/debriefing	0.642	Substantial
	Clinical guidelines	0.196	Slight
	Accessible medical records	0.493	Moderate
Cost-effectiveness	Financial transparency	0.548	Moderate
	Community cost–benefit analysis	0.075	Slight

The STAT comprises one stage of a clinical audit cycle for organizations delivering STMMs.[26] While problem identification and standard setting are comprehensively addressed in the previous content validation paper,[13] the present study addresses one mechanism for rapid, binary data collection and comparison of performance with stakeholder-validated criteria. It remains to be seen whether the tool will have utility in the final stage of the cycle, which involves implementing change. In sum, the STAT may be useful in the context of an internal review of STMM policies.

Assessment of the inter-rater reliability of the STAT is essential, since a key challenge to such projects is the tendency for bias among assessors who are intimately involved with an organization.[8] Nonetheless, in the absence of a supervening governing body to independently certify such organizations, self-assessment remains the most feasible mechanism for quality improvement for the immediate future. Self-regulation of healthcare professionals with government oversight is already a common and well-defined approach in professions such as medicine and nursing across high-income countries. This quality assurance often involves intentional self-reflection as an integral part of a process of identifying strengths, areas for improvement and learning needs.

A validated online tool would allow prospective volunteers and donors to make decisions based on the positive and negative reviews of previous volunteers, thus shifting demand toward high quality projects. Likewise, when provided with a validated framework for responsible STMM projects, organizations themselves may be encouraged to adapt their practices to meet the evolving expectations of volunteers, hosts and the global health community at large. As such, the STAT may be deployed by NGOs as a self-reflection exercise in order to examine strengths and weaknesses, by host organizations to hold their partners accountable and highlight gaps, or by governments in order to allow cross-comparisons of the organizations operating within their borders.

Strengths

The STAT builds on the existing literature, including the most recent systematic review on the topic of STMM best practices,[6] our initial eDelphi discussion-based framework for best practices,[13] and is structured based on the only other published quantitative assessment tool for STMMs.[8] To our knowledge, this is the largest existing quantitative study of STMM best practices and the first to attempt a data-driven assessment of practice quality. This pilot study includes a broad sample of STMMs, compiled through the integration of multiple databases, and involved multiple attempts to contact each NGO through multiple channels, including social media. Furthermore, each element included in the STAT tool is theoretically falsifiable, making this tool more objective than previously developed assessment tools.

Limitations

This pilot study has several limitations. First, despite multiple attempts at contact, 70% of STMM organizations were non-responders. While this response rate exceeds that of typical external e-mail surveys,[20,21] it nonetheless suggests an important response bias. One might speculate that certain organizational characteristics make an NGO more likely to participate in quality improvement initiatives, such as its size and resources, funding sources, overall philosophical mission or other factors yet to be determined. However, while the views and characteristics of

these non-responders may not be represented at present, the evolving, public and online nature of the tool makes it possible for such groups to become involved during the external validation phase. Second, this pilot was administered in English only, which limits participation from predominantly Spanish-speaking host country administrators in LAC.

Third, it is important to note that the 18 items in the STAT inventory represent only those items that achieved consensus in an eDelphi process. While these items were agreed to be essential to a high quality STMM, the list is not exhaustive and does not negate the potential importance of items that did not achieve consensus. In theory, the absence of a fulsome assessment of these missing elements could impart a false sense of confidence and reassurance to volunteers in their selection of an STMM experience. Therefore it is more accurate to suggest that while it is problematic if an organization does not incorporate the STAT elements, volunteers should continue to perform their due diligence even in cases where the STMM appears to perform well. Volunteers should carry out conscientious vetting, even for organizations that fulfil all 18 criteria, to ensure that all legal and ethical requirements are being met.

Finally, the STAT also makes no comment on the relative importance of the elements in the inventory. It is conceivable that for many volunteers, the presence of a local clinician, a permanent presence in the community and an adequate system of referral would be more convincing indicators of quality than transparency of finances or the presence of a system of scheduling and triage. However, until a rational, quantitative mechanism can be elucidated to weigh the individual components of the STAT, their relative importance must continue to be dictated by the values and preferences of those evaluating each opportunity. It is therefore an ordinal scale; as such, a higher score suggests a higher quality medical mission but does not necessarily correlate with any given degree of improvement.

Future directions

Additional improvements to the tool are necessary to ensure that it accurately reflects the quality of STMMs. Most critically, current work aims to integrate host community voices to expand the framework. Refinement of the phrasing of STAT questions that performed poorly in this pilot, as well as broader external validation with larger sample sizes and in diverse settings, are also necessary to support further refinement of the tool. Since this study reports only the claimed adherence of NGO administrators and volunteers to best practices, further validation by fully independent and standardized assessors would be valuable to correlate these reports to verifiable indicators.

While we can speculate that volunteers and charitable organizations themselves are predominantly motivated by a desire to do good, other competing motivations may be less altruistic.[27,28] As such, any quality improvement exercise must consider whether enforcement is necessary for those STMMs that cannot be engaged through traditional means. While we believe the data produced by the STAT project would be helpful in this regard, the legal implications of enforcement are outside the scope of this study, although this topic is discussed in depth in a recent publication.[9] It is also reasonable to believe that a similar tool could be useful for other, non-medical cadres, although such a tool would need to be separately validated.

Conclusions

The STAT tool presents a novel quantitative approach for assessing STMMs in developing countries and may be a viable resource for volunteers who are considering participation, NGOs seeking to improve their practices, donors looking to make financial or in-kind contributions to support charitable work and governments tasked with the regulation of their healthcare workforce.

Authors' contributions: CD conceived the study and designed the study protocol, conducted statistical analysis and drafted the manuscript. CG contacted and conducted follow-up with STMM sending organizations. All the authors critically revised the manuscript for intellectual content and read and approved the final manuscript. CD is guarantor of the paper.

Acknowledgements: None.

Funding: None.

Competing interests: None declared.

References

1 Sykes K. Short-term medical service trips: a systematic review of the evidence. Am J Public Health. 2014;104(7):e38–e48.

2 Caldron PH, Impens A, Pavlova M, Groot W. A systematic review of social, economic and diplomatic aspects of short-term medical missions. BMC Health Serv Res. 2015;15:380.

3 DeCamp M, Lehmann LS, Jaeel P, Horwitch C. Ethical obligations regarding short-term global health clinical experiences: a An American College of Physicians Position Paper. Ann Intern Med. 2018;168(9):651–57.

4 Shah S, Lin HC, Loh L. A comprehensive framework to optimize short-term experiences in Global Health (STEGH). Global Health. 2019;15:27.

5 Norton I, Schreeb J, Aitken P, Herard P, Lajolo C. Classification and minimum standards for foreign medical teams in sudden onset disasters. Geneva: World Health Organization, 2013. Available from: https://www.who.int/csr/resources/publications/ebola/foreign-medical-teams/en/ [accessed 27 July 2019].

6 Roche SD, Ketheeswaran P, Wirtz VJ. International short-term medical missions: a systematic review of recommended practices. Int J Public Health. 2017;62(1):31–42.

7 Lasker JN. Hoping to help: the promises and pitfalls of global health volunteering. Ithaca, NY: Cornell University Press, 2016.

8 Maki J, Qualls M, White B, Kleefield S, Crone R. Health impact assessment and short-term medical missions: a methods study to evaluate quality of care. BMC Health Serv Res. 2008;8:121.

9 Rowthorn V, Loh L, Evert J, Chung E, Lasker J. Not above the law: a legal and ethical analysis of short-term experiences in global health. Ann Global Health. 2019;85(1):79.

10 Seager G. When healthcare hurts: an evidence based guide for best practices in global health initiatives. Bloomington, IN: AuthorHouse,

2012.

11 Caldron PH, Impens A, Pavlova M, Groot W. Economic assessment of US physician participation in short-term medical missions. Global Health. 2016;12(1):45.

12 Lasker JN, Aldrink M, Balasubramaniam R, et al. Guidelines for responsible short-term global health activities: developing common principles. Global Health. 2018;14:18.

13 Dainton C, Gorman C, Chu CH, Cherniak W. Development of a theoretical framework for assessment of quality of primary care medical service trips (MSTs) in Latin America. Int J Public Health. 2019;64:333–42.

14 Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: Development and pilot validation. Syst Rev. 2014;3:82.

15 Ray KN, Miller E. Strengthening stakeholder-engaged research and research on stakeholder engagement. J Comp Eff Res. 2017;6(4):375–89.

16 Crump JA, Sugarman J. Ethics and best practice guidelines for training experiences in global health. Am J Trop Med Hyg. 2010;83(6):1178–82.

17 Dillman DA, Smyth JD, Christian LM. Internet, phone, mail, and mixed-mode surveys: the tailored design method. Hoboken, NJ: John Wiley & Sons, 2014.

18 Wells CS, Wollack JA. An instructor's guide to understanding test reliability. Madison, WI: Testing and Evaluation Services, University of Wisconsin, 2003.

19 Dainton C, Shah, Chu CH. Prevalence of portable point of care tests used on medical service trips in Latin America and the Caribbean. Ann Global Health. 2018;84(4):736–42.

20 FluidSurveys Team. Response rate statistics for online surveys – What numbers should you be aiming for? Available from: http://fluidsurveys.com/university/response-rate-statistics-online-surveys-aiming/ [accessed 9 January 2019].

21 Fryrear A. What's a good survey response rate? Available from: https://www.surveygizmo.com/resources/blog/survey-response-rates/ [accessed 5 February 2019].

22 Randolph J. Online kappa calculator. Available from: http://justus.randolph.name/kappa [accessed 24 September 2019].

23 McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276–82.

24 Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971;76(5):378–82.

25 Brennan RL, Prediger DJ. Coefficient kappa: some uses, misuses, and alternatives. Educ Psychol Meas. 1981;41(3):687–699.

26 Benjamin A. Audit: how to do it in practice. BMJ. 2008;336(7655):1241–45.

27 Caldron PH, Impens A, Pavlova M, Groot W. Why do they care? Narratives of physician volunteers on motivations for participation in short-term medical missions abroad. Int J Health Plan Manage. 2017;33(1):67–87.

28 Rovers J, Japs K, Truong E, Shah Y. Motivations, barriers and ethical understandings of healthcare student volunteers on a medical service trip: A mixed methods study. BMC Med Educ. 2016;16:94.

APPENDIX A

Service Trip Audit Tool (STAT)
The STAT tool is designed to assess the quality of short-term medical missions.

Domain	Minor elements	YES	NO	NOT SURE
Sustainability	1. The organization has a formal partnership with local health services in the host community.	☐	☐	☐
	2. The organization has a clear referral process for patients who need higher levels of care.	☐	☐	☐
	3. In addition to the visiting volunteers, the organization ensures that there is always a local clinician involved in clinical care	☐	☐	☐
	4. The organization has a permanent staff member or partner organization in the host community	☐	☐	☐
Comments:				
Education	1. The organization builds capacity by helping train host providers, local health workers, or community health workers.	☐	☐	☐
	2. The organization engages in public health work or health promotion in the community.	☐	☐	☐
Comments:				
Efficiency	1. The organization promotes the visiting clinics to locals by word of mouth or advertisement, or uses a clinic location that is already well known to locals.	☐	☐	☐
	2. The organization has a formal staffing plan describing future needs and a recruitment strategy.	☐	☐	☐
	3. The organization has a formal triage, priority, appointment, or ticketing system in place for patients visiting the clinic.	☐	☐	☐
Comments:				
Impact and safety	1. The organization solicits written feedback/debriefing from volunteers after the trip is over.	☐	☐	☐
	2. The organization keeps medical records that are easily accessible to future clinicians.	☐	☐	☐
	3. The organization provides evidence-based clinical guidelines to volunteers, describing an approach to common diseases in the host community.	☐	☐	☐
Comments:				
Preparedness	1. Volunteers are pre-screened before being accepted by the organization.	☐	☐	☐
	2. The organization provides pre-departure training for volunteers (i.e. in-person or online).	☐	☐	☐
	3. Urine dipsticks, pregnancy tests, and glucometers are all available, and there is a clear pathway for volunteers to obtain more advanced tests.	☐	☐	☐
	4. The organization provides written clinical protocols to volunteers (i.e. limiting their practice scope to the care they are licenaed to provide at home)	☐	☐	☐
Comments:				
Cost effectiveness	1. The financial reports for this organization are transparent and easily available (i.e. via website, annual report, etc.)	☐	☐	☐
	2. The organization considers and describes any host community costs that are associated with hosting volunteers (i.e. on their website).	☐	☐	☐
Comments:				