BMJ Global Health

Third party monitoring for health in Afghanistan: the good, the bad and the ugly

Sandra Alba ,[1] Eelco Jacobs,[1] Elisabeth Kleipool,[1] Ahmad Salehi,[2] Ahmad Naeem,[3] Sayed Rahim Arab,[4] Margo Van Gurp,[1] Nasir Hamid,[5] Partamin Manalai,[6] Sayed Ataullah Saeedzai,[7] Sohrab Safi,[6] Farhad Paiman,[8] Abdul Majeed Siddiqi,[9] Barend Gerretsen,[1] Sara Gari,[10] Egbert Sondorp [1]

To cite: Alba S, Jacobs E, Kleipool E, *et al*. Third party monitoring for health in Afghanistan: the good, the bad and the ugly. *BMJ Glob Health* 2023;**8**:e013470. doi:10.1136/bmjgh-2023-013470

Handling editor Seye Abimbola

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjgh-2023-013470).

Received 19 July 2023
Accepted 1 November 2023

For numbered affiliations see end of article.

Correspondence to
Dr Sandra Alba; s.alba@kit.nl

ABSTRACT

Third party monitoring (TPM) is used in development programming to assess deliverables in a contract relationship between purchasers (donors or government) and providers (non-governmental organisations or non-state entities). In this paper, we draw from our experience as public health professionals involved in implementing and monitoring the Basic Package of Health Services (BPHS) and the Essential Package of Hospital Services (EPHS) as part of the SEHAT and Sehatmandi programs in Afghanistan between 2013 and 2021. We analyse our own TPM experience through the lens of the three parties involved: the Ministry of Public Health; the service providers implementing the BPHS/EPHS; and the TPM agency responsible for monitoring the implementation. Despite the highly challenging and fragile context, our findings suggest that the consistent investments and strategic vision of donor programmes in Afghanistan over the past decades have led to a functioning and robust system to monitor the BPHS/EPHS implementation in Afghanistan. To maximise the efficiency, effectiveness and impact of this system, it is important to promote local ownership and use of the data, to balance the need for comprehensive information with the risk of jamming processes, and to address political economy dynamics in pay-for-performance schemes. Our findings are likely to be emblematic of TPM issues in other sectors and other fragile and conflicted affected settings and offer a range of lessons learnt to inform the implementation of TPM schemes.

SUMMARY BOX

- ⇒ In this paper, we report and analyse the perceptions of selected stakeholders involved in the third party monitoring (TPM) of two subsequent programs that implemented the Basic Package of Health Services (BPHS) and the Essential Package of Hospital Services (EPHS) in Afghanistan between 2013 and 2021.
- ⇒ *The good:* Ministry of Public Health (MOPH) representatives perceived TPM as a system that provided reliable data to support evidence-based decision-making at a reasonable cost (proportional to overall SEHAT/Sehatmandi implementation costs).
- ⇒ *The bad:* In practice, in MOPH and TPM representatives' perception, TPM implementation was hampered by the overwhelming amount of data collected, analysed and reported for TPM purposes. Furthermore, service providers' representatives voiced substantial discontent regarding TPM implementation after the introduction of the pay-for-performance mechanism in 2018.
- ⇒ *The ugly:* Several accusations of data fabrication, corruption and undue influence related to the TPM implementation were reported to have circulated. While these accusations cannot be substantiated, they shed light on the interests at stake and the perverse incentives associated with TPM activities, especially when they are linked to pay-for-performance mechanisms.
- ⇒ The TPM shortcomings we experienced may be emblematic of TPM in other settings and could be mitigated by promoting local ownership and use of the data, balancing the need for comprehensive information with the risk of jamming processes, and addressing political economy dynamics in pay-for-performance schemes.

INTRODUCTION

Good and evil are opposite ends of a single spectrum, and there is a range of possibilities between them. […] It is the task of the wise person to discern what is truly good and to avoid what is truly evil. Abu Ali Ibn Sīna—The Book of Healing (AD 1027)

Third party monitoring (TPM) is increasingly used in development programming. It may fulfil various roles, ranging from acting as the ears and eyes of donors for activities implemented in insecure areas to independent performance verification in the context of a contract between two parties.[1] The latter includes the assessment of deliverables in a contract between a purchaser (usually donor or government) and providers (non-governmental organisations (NGOs)

Box 1 SEHAT and Sehatmandi programmes

After the removal of the first Taliban regimen in 2001, the establishment of a new government and promises of generous foreign aid paved the way for new avenues to address Afghanistan's extreme healthcare challenges. Prior to 2001,[16] for many years, approximately 70% of the country's limited healthcare services were delivered by roughly 20 international and national non-governmental organisations.[13] Building on this, a novel health policy and strategy emerged, with the Ministry of Public Health (MoPH) providing stewardship over the health sector and contracting non-state entities as the primary providers of newly established essential health packages, the Basic Package of Health Services (BPHS) and the Essential Package of Hospital Services (EPHS). This contracting model was rolled out across the country with the purpose to reach all Afghans.[23] The new national health strategy, with delivery of the EPHS and BPHS packages as backbone, was supported by three key donor agencies - World Bank, United States Agency for International Development (USAID) and European Union - initially each with different institutional arrangements. Over time approaches were aligned. Funds were pooled and increasingly managed by the newly created Grants and Contracts Management Unit within the MoPH. These arrangements were completed during the 2013–mid-2018 health project called SEHAT.[18] The successor project, Sehatmandi, was planned for mid-2018–2022[24] but financial support ended abruptly in August 2021 when the Taliban took control over the country.[19] Despite the highly challenging and fragile context, the large investments in the implementation of the BPHS/EPHS (including the use of the contracting-out method) have been acknowledged as important factors contributing to the remarkable advancements in Afghanistan's public health in the two decades since 2001.[25–27]

or other non-state entities). This attention to measuring outcomes and results is in line with the aid effectiveness paradigm[2] and new public management philosophy[3] that gained traction in health service delivery since the 1990's. Afghanistan's health sector has known TPM since 2003 when contractual arrangements started to emerge between the Ministry of Public Health (MoPH) and NGOs as service providers (SPs). However, not much is (publicly) known about the processes, dynamics and challenges of Afghanistan's health sector TPM. Here, we present a unique reflection from the three parties involved in the TPM of the Afghan health sector between 2013 and 2021. More specifically, we draw from our experience as public health professionals involved in the SEHAT and Sehatmandi programmes that implemented two essential health packages—the Basic Package of Health Services (BPHS) and the Essential Package of Hospital Services (EPHS)—between 2013 and 2021 (see box 1 for details). We first provide contextual information about the TPM activities, then analyse our own TPM experience through the lens of each of the three parties involved. We conclude with insights to inform future health sector TPM approaches in Afghanistan and in other fragile and conflict affected settings.

From the start, independent TPM agencies have monitored the contracting of the BPHS/EPHS in Afghanistan. Up to 2013, the main instrument was a nationwide annual health facility survey to measure service delivery and quality, summarised in an overall Balanced Scorecard index for each province.[4 5] During the SEHAT project (starting in 2013), the annual Balanced Scorecard surveys continued, and a new tool was added, the Health Management Information System (HMIS) Verification and Health Facility Functionality Assessments. A portion of the payments to SPs (up to 20% of the contract value) became linked to the results of the bi-annual HMIS Verification scores.[6] With the Sehatmandi program (starting in 2018), a new pay-for performance model was introduced whereby payments became much more dependent on the volume of service delivery and the frequency of reports increased from bi-annual to quarterly.[6] In addition, the TPM included several other assessments that did not affect payments, such as the drug quality assessments and nationwide household surveys.[7] Under both SEHAT and Sehatmandi, KIT Royal Tropical Institute, an independent knowledge institute based in the Netherlands, has been involved as independent TPM. KIT initially collaborated with the Afghan not-for-profit organisation Silk Route Training and Research Organisation (under SEHAT) and later with the independent consultancy company Particip, based in Germany (under Sehatmandi).

While the implementation of contracting systems for heath service delivery in fragile and conflict affected settings has been the subject of intense scholarly debate,[8–11] associated TPM processes, dynamics and challenges have not been documented. To the best of our knowledge, there is currently no analysis of the experiences and perceptions of the three main parties involved in a single TPM scheme, at the same time. How do the three parties perceive each other? Do dynamics change over time? Are any unintended effects set in motion by these relations? In the case of Afghanistan, the three parties in question are: (1) the Ministry of Public Health (MoPH) as contractor of health SPs, responsible for setting standards for healthcare delivery and ensuring that these are met; (2) the NGOs as SPs contracted to deliver the BPHS/EPHS; (3) the TPM agency, collecting and analysing data to monitor the implementation of the BPHS/EPHS and reporting to MoPH on the extent to which targets and standards are met.

With essential packages increasingly promoted as a means to achieve universal health coverage[12] (and implementation contracted-out in fragile and conflict affected settings[13 14]) development practice[15] calls for specific attention to their monitoring processes. This is especially important for TPM mechanisms since they operate at the junction of various power structures.[1] Reflexivity is the process of being able to examine one's positionality, and how this influences what one thinks.[16 17] Attention to power relations and stakeholder interests is an integral part of this. In the case of Afghanistan's health TPM, TPM staff inevitably brought their own assumptions into all steps of the TPM process, while SPs and MoPH staff made meaning of the TPM process from their own identities

and experiences. Reflecting on this requires all parties to exercise dynamic self-awareness in order to deconstruct their positionality and produce a trustworthy, transparent and honest account of the TPM enterprise. Reflexive practice is a methodological self-consciousness that has expanded from anthropology into mainstream of qualitative research, including but not limited to the field of global health. In line with these developments, we believe this TPM reflection is a timely and much needed exercise in order to better understand, position and ultimately improve TPM operations in Afghanistan and beyond.

To gather the views of the TPM parties on the TPM experience during the SEHAT and Sehatmandi years, KIT advisors involved in both TPM programmes (see box 1) organised a series of consultations. Fifteen representatives of the three parties were invited to share their views in either round table discussions or one-to-one interviews and were invited to contribute to this publication (MoPH: n=4, SPs: n=5 from 4 different NGOs, TPM: n=6). Eleven explicitly agreed to become co-authors (MoPH: n=3; SPs: n=3; TPM: n=5). Topic guides encouraged participants' reflections on the five criteria for evaluations defined by the Organisation for Economic Co-operation and Development (OECD): relevance, efficiency, effectiveness, impact and sustainability. We chose the OECD criteria as they provide a common language to appraise development interventions. Guiding questions were developed for each OECD criterion and each party, in relation to SEHAT and Sehatmandi's TPM objectives (online supplemental file 1). As per the MoPH tender documents, the TPM objectives for both programmes can be summarised as follows: to verify service delivery, to assess the performance and progress of the health SPs and to identify the strengths and weaknesses of the BPHS and EPHS implementation financed by SEHAT/Sehatmandi. We synthesised the three parties' views using the popular headers 'the good, the bad and the ugly' to provide an overview of the positive intended effects (the good), the negative side effects (the bad), and the negative incentives (the ugly). Despite our intentions to offer an honest reflection from the perspective of the three parties, this practice report is inevitably subject to potential biases resulting from the authors' positionality and some methodological limitations. In box 2, we give consideration to these issues and describe our mitigation measures in the processes of data collection, synthesis and reporting.

Box 2 Authors' positionality, methodological limitations and mitigation measures

⇒ Participants from the Ministry of Public Health (MoPH), or formerly associated with the MoPH, were high level senior staff with broad overview on all aspects of the health system, including on the role that third party monitoring (TPM) data played at senior managerial level within the MoPH. Service provider (SP) participants worked in various non-governmental organisation management positions and were based in Afghanistan during the SEHAT/Sehatmandi years; TPM participants were involved in TPM operations at several levels (project management, field operations and data analysis) and were based in either Afghanistan, the Netherlands or Germany.

⇒ TPM staff based in the Netherlands initiated and facilitated the round table discussions and one-on-one interviews and took the lead in developing the first draft of the manuscript. This lead position gave TPM a certain degree of power in framing the message of the manuscript.

⇒ For pragmatic reasons, the input from the various consultations could not be synthesised with active participation of all people involved. However, we have taken care to offer multiple, and sometimes conflicting perspectives on the emerging challenges and bottlenecks, with attention to the underlying political economy and resulting power dynamics.

⇒ The author team—composed of representatives from all the three parties—went through iterative rounds of feedback, adaptation, rewriting and reframing to allow for the inclusion of all voices, resulting in a manuscript on which all authors agreed.

THE GOOD

Overall MoPH and TPM representatives perceived the TPM as a system that provided reliable data to support evidence-based decision.

From its inception, MoPH staff saw the importance and relevance of the TPM as part of the contracting out model, to ensure a separation of the functions of stewardship (by MoPH), service provision (by both national and international NGOs) and monitoring and verification (by an independent TPM). Within this design, TPM provided data to support evidence-informed decision-making and to ensure accountability towards various constituencies: 'upwards' towards MoPH and donors (TPM submitted reports and databases to MoPH) and 'downwards' towards the Afghan population (TPM reports were uploaded on the MoPH website for public consultation). By and large, MoPH representatives considered TPM an efficient and worthy investment with costs proportional to purpose (less than 2% of SEHAT[18] and Sehatmandi[19] grants were earmarked for TPM).

TPM and MoPH representatives generally agreed that TPM products were effective in the sense that they fulfilled their primary purpose of providing evidence (to MoPH and donors) on health service delivery. In the process, the need to provide TPM with unambiguous cut-offs and decision-making rules led to the definition of clear minimums standards of services such as those defined in the Sehatmandi Standard Operating Procedures. TPM and SP staff also noted a positive effect of HMIS verification exercices, as they improved the consistency and accuracy of HMIS reports over time, with a positive overall impact on the country's health information system. Some SP representatives appreciated the insights TPM reports provided on infrastructure and availability of equipment, stating that TPM data also helped bring improvements (as TPM data could be used for advocacy and to back funding requests). In line with this, according to MoPH representatives, TPM reports were used by the Policy and Planning department to develop tailor-made

improvement plans together with SPs. More specifically, one MoPH representative provided an example regarding the use of Balanced Scorecard reports to provide SPs with concrete action plans for performance improvement, by highlighting existing gaps in the availability of staff and equipment in health facilities. This view was echoed by TPM staff who perceived an increased evidence-based culture within MoPH, reflected by multiple requests from different MoPH units to be trained on (TPM) data analysis and interpretation.

Most MoPH, TPM and SP representatives agreed that reflecting on the sustainability of a highly donor-dependent system such as Afghanistan's BPHS/EPHS contracting model is hypothetical at best. However, overall, they agreed that the TPM provided key monitoring and accountability functions to guarantee continued funding. Following the logic of evidence-based development practice, evidence of continued service delivery in Afghanistan could have provided assurance to donors that funds were being appropriately channelled, thereby contributing to their willingness to continue supporting the health system over past decades.

Table 1 Examples of TPM data verification procedures

Party	TPM data verification procedures
MoPH	▶ Review of TPM final reports and triangulation between TPM and non-TPM data sources. ▶ Independent resurvey to verify TPM data in a sample of the health facilities surveyed by TPM, using the same data collection tools for the same reporting period (done once in 2021, overall no significant difference was found between the two data sources).
SPs	▶ Verification TPM (draft and final) data against health facility registers. ▶ Rechecking availability of medicines, human resources and equipment, etc in health facilities.
TPM	▶ Geo-tagged photos of health facilities to ensure that all were physically reached by monitors. ▶ Pictures and scans of registers of source data collected from health facilities. ▶ Signing off of TPM data collection in health facilities by health facilities in-charge as proof that collected data matched registers. ▶ Recollecting data in 10% of facilities by independent monitors. ▶ Review of statistical programming files by independent analysts. ▶ Sharing draft results with SPs to enable review and potential justified data corrections.

MOPH, Ministry of Public Health; SPs, service providers; TPM, third party monitoring.

THE BAD

Despite several positive achievements, all parties—MoPH, SPs and TPM—agreed that in practice, the TPM was hampered by several limitations inherent to its design and implementation.

SP representatives reported a gradual shift in the main function of the TPM from SEHAT to Sehatmandi - from what they perceived as a potentially *relevant* tool to inform health service delivery ('learning' function of monitoring and evaluation) during the SEHAT years to a mechanism primarily aimed at triggering payments ('accountability' function) in the Sehatmandi years. This coincided with two important developments: the involvement of more national NGOs as SPs; and the introduction of a new pay-for-performance service delivery model which paid SPs based on the volume of key services they provided.[6] In this pay-for-performance model, the role of the TPM was to verify that the volume of services reported by SPs was accurate and to prevent potential misreporting incentivised by linking service volumes to payment sums.

According to SP representatives, these developments were responsible for introducing several inefficiencies in the TPM. Indeed, extensive verification mechanisms were put in place to vouch for the veracity of TPM data when it became directly tied to performance-based payments (table 1). TPM representatives underscored that these extensive mechanisms were warranted as an added layer of security because of the involvement of more national NGOs, that increased the fear that personal relationships between individuals from SPs, MoPH and TPM could lead to biased TPM data (and payments) in favour of certain NGOs. However, this intensive scrutiny was perceived by SPs to have contributed to long payment delays, with cascading effects for healthcare delivery. Furthermore, MoPH did not allow direct communications between SPs and TPM during report compilation and finalisation which added to the delays (since all communications were centralised through the MoPH).

In addition, MoPH, SP and TPM representatives agreed that the implementation of the TPM was not as effective as it could have been. SP and TPM representatives lamented the frequency, number and size of the surveys conducted. As a result, SPs experienced TPM data collection as intrusive, placing a burden on their capacity for service delivery. Furthermore, SPs perceived TPM reports as lacking both cohesion and consistency—because there were too many indicators and because similar indicators were sometimes presented in different reports with apparently contradicting results. Indeed, there was some overlap in the indicators of the Balanced Scorecard and the Health Facility Functionality Assessments, particularly when it comes to availability of medicines, human resources and equipment. These indicators were not measured in a similar manner in both assessments and as such produced different outcomes in their respective reports, causing confusion among SPs. In addition, SPs voiced concerns about what they perceived as inadequate training of TPM data collectors and the burden this placed on them as they felt compelled to shadow TPM staff during collection (since incorrect data could have serious negative consequences). TPM staff acknowledged that training data collectors was challenging, partly due to the considerable number of surveys and quick reporting turnaround. These views were shared by MoPH representatives who conceded the enormity and difficulty of the TPM endeavour, but also agreed with SPs that TPM technical capacity on the ground was not up to standard

especially in SEHAT years, attributing it to the limited presence of TPM international staff in Afghanistan.

Further challenges related to effectiveness were voiced by TPM and MoPH representatives, pertaining to the complexity of the TPM design, especially during the Sehatmandi years. Indeed, the transition from SEHAT to Sehatmandi saw the introduction of both pay-for-performance indicators and minimum standards of services. Consequently, the Sehatmandi TPM verification tools had two different objectives: (1) the HMIS verification aimed to verify the volume of service delivery for the eleven pay-for-performance indicators, which were linked to payments; and (2) the functionality verification aimed to verify the performance of the SPs towards the achievement of the minimum standards of services as per the BPHS and EPHS guidelines. Both measurements contributed to the overall performance scores of the SPs (used to assess contracts, including recognition, rewards, contract extension, warning letters, performance improvement plans, termination, etc) but only HMIS verification was linked to payments. However, according to TPM representatives these technical subtleties were not always clear to MoPH and SP staff. TPM results were thus often misunderstood and misinterpreted. This hampered the usefulness of TPM data beyond purely contract management purposes: opportunities were missed to provide SPs with constructive feedback to improve performance other than penalising them by withholding payments. To address this issue, from 2019 to 2021, a Performance Management Office was set up within the MoPH, with the purpose of reviewing data with the SPs and support use of data for performance improvement. Nevertheless, MoPH representatives echoed similar concerns about underutilisation of TPM data for performance improvement and programme planning, while also citing difficulties of making sense of the overwhelming amount of data and highly technical nature of the reports. One MoPH representative attributed this to siloed structures within the MoPH that separated departments with strong analytical capacities from departments with programming mandates (eg, provincial health authorities and MoPH's department of planning).

THE UGLY

In addition to design and implementation issues, MoPH, SP and TPM representatives reported the circulation of several (unsubstantiated) accusations of data fabrication, corruption and undue influence related to the TPM implementation.

SPs and MoPH representatives' main concern was data fabrication by TPM staff when they were supposedly unable or unwilling to travel to parts of the country, because of either security issues (checkpoints and kidnappings by Taliban or other armed, sometimes criminal, groups) or geographical accessibility (Afghanistan being a highly mountainous country with poor road infrastructure and snowy winters). On one hand TPM representatives adamantly defended the good faith of the TPM endeavour. They referred to the extensive mechanisms put in place during Sehatmandi to verify each of the accusations and to provide proofs and corrective actions if needed. On the other hand, TPM staff also conceded the risk of professionalising monitoring and evaluation with a fixed pool of data collectors who in some cases also professionalise their ability to cut corners. As a result, quality assurance processes needed to become increasingly complex and tended to focus more on corrective actions (responding to issues as they arose) rather than preventive strategies (anticipating issues in the field).[20]

Scratching under the surface of these concerns over data quality revealed a complex network of personal relationships and vested interests. Indeed, triangulating accounts from MoPH, SP and TPM representatives, five main types of rumours were identified, involving some Taliban forces and other armed groups, some SPs, some TPM data collectors and some individuals within the MoPH. First, some TPM data collectors allegedly exaggerated security challenges to produce timely and seemingly complete reports (bar those health facilities claimed to be unreachable) and fulfil contractual obligations. Second, some SPs seemingly instrumentalised security challenges as an excuse to block access to some badly performing HFs or to intimidate TPM data collectors (intimidation by armed groups was orchestrated by SPs and blamed on the Taliban). Third, some SPs presumably influenced certain TPM data collection teams to produce favourable scores (either with outright payments or indirectly by providing transportation, food and accommodation). Fourth, certain MoPH staff supposedly applied pressure to fire TPM staff not complying with collaborative requests by SPs (linked to certain MoPH individuals) based on accusations of data fabrication. Fifth, some individuals from MoPH reportedly pressurised TPM to adjust the scores of certain SPs (linked to certain MoPH individuals) based on data fabrication arguments when scores were too low. While the verification and rebuttal of these accusations is beyond the scope of this article, we believe that reporting them is worthwhile as they were widely shared - and people likely acted on them whether they were true or not. In other words, these rumours make explicit some of the stakes at hand and highlight the interplay between technical and political dimensions of TPM work.

Against this backdrop, it is important to note that SPs do not represent one homogeneous entity. Two main differences are most relevant here: (1) national (Afghan) versus international NGOs and (2) size and availability of contingency funds beyond SEHAT/Sehatmandi contracts to withstand payment delays. On one hand suspicions of personal links between individuals working in SP organisations and individuals at MoPH appear to be mostly targeted at large national NGOs with considerable clout in government circles. On the other hand, some SP representatives were adamant that they were not in the financial position to spare funds to influence TPM data collectors and that they did not have the political

motivation to influence Taliban forces. This highlights the instrumental role not only of financial, but also social capital, in implementing service provision contracts noted in other fragile and conflict-affected settings.[21] These SPs reported that their dealings with Taliban or other armed forces were limited to negotiating access to their facilities for patients, as well as undergoing increasingly strict parallel health facility monitoring by Taliban public health officers. For smaller SPs entirely dependent on SEHAT/Sehatmandi contracts, payment delays resulting from accusations of data fabrication led to vicious circles of low cash-flow affecting service delivery and subsequent TPM reports and therefore future access to funds.

CONCLUSIONS

Despite the highly challenging and fragile context, the consistent investments and strategic vision of donor programmes in Afghanistan over the past decades have led to a robust monitoring system to support the implementation of the BPHS/EPHS. TPM has shown its relevance as a model first intended to monitor the functioning of the BPHS/EPHS and later to verify service delivery for accountability purposes.

Based on SP, MoPH and TPM reflections, there are several ways to maximise the performance of a TPM in a fragile and conflict affected setting such as Afghanistan. While SPs voiced most of the criticism against the TPM, they also experienced most of the repercussions (delayed or withheld payments). Not only could MoPH and TPM staff often relate to SP concerns, they also had several of their own. Consolidating all views, three main recommendations emerge to related to the efficiency, effectiveness and impact of a TPM. First, to ensure impact, mechanisms need to be put in place to foster TPM data use by local actors, as this can foster trust, understanding and local ownership of the system. Second, for both effectiveness' and efficiency's sake, the need for a comprehensive information system needs to be balanced with the risks of over-burdening (almost to the point of jamming) processes—as exemplified by the challenges reported with complex pay-for-performance algorithms, overwhelmed data systems, intrusive data collection processes, and long delays in verification/payments. Finally, when TPM activities are linked to pay-for-performance schemes, it is important that all parties involved are attuned to the conflicts of interests and perverse incentives at play—as testified by the accusations of data fabrication, corruption and undue influence that were reported. Overlooking such important political economy dynamics in the long run can threaten not only the efficiency, effectiveness and impact of a TPM, but also its very relevance.

Afghan TPM experiences in health have encompassed the whole spectrum between good and ugly, likely emblematic of TPM experiences in other sectors and other fragile and conflicted affected settings.[22] In the words of the Persian physician and philosopher Ibn Sina, our reflections offer a range of possibilities for donors and implementers to discern a wise implementation of TPM schemes.

Author affiliations
[1]KIT Royal Tropical Institute, Amsterdam, The Netherlands
[2]Global SienCentre, Toronto, Ontario, Canada
[3]Assistance for Families and Indigent Afghans to Thrive (AFIAT), Kabul, Afghanistan
[4]Particip, Kabul, Afghanistan
[5]Care of Afghan Families (CAF), Kabul, Afghanistan
[6](formerly) Particip, Kabul, Afghanistan
[7](formerly) M&E HIS, Ministry of Public Health, Kabul, Afghanistan
[8]Organization for Health Promotion and Management, Kabul, Afghanistan
[9]HealthNet TPO, Amsterdam, Netherlands
[10]Particip, Freiburg, Germany

Twitter Sandra Alba @san_dra_al_ba

Acknowledgements We are very grateful to Johannes Ohnmacht, Project Director of Sehatmandi TPM (Particip) for facilitating the development of this manuscript and providing crucial input.

Contributors SA, EJ, EK, BG and ES conceived the manuscript. SA, EJ and EK initiated and led consultations with the other coauthors. SA wrote the first draft of the manuscript. All authors revised it critically and provided important intellectual content. All authors provided substantial contributions to interpretation of data. All authors agree to be accountable for all aspects of the work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer The experiences described and recommendations expressed in this manuscript are those of the authors alone and do not necessarily reflect the views or official policies of our organisations or funders.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

ORCID iDs
Sandra Alba http://orcid.org/0000-0003-2435-624X
Egbert Sondorp http://orcid.org/0000-0003-2325-7266

REFERENCES

1 Harrison MR. *Study on best practices in third party monitoring*. Brussels: International Consulting Expertise on behalf of the European Commission, August 2020. Available: https://fpi.ec.europa.eu/system/files/2021-05/study_on_best_practices_in_third_party_monitoring__0.pdf
2 Dalil S, Newbrander W, Loevinsohn B, *et al*. Aid effectiveness in rebuilding the Afghan health system: a reflection. *Glob Public Health* 2014;9 Suppl 1:S124–36.
3 van de Walle S. New public management: restoring the public trust through creating distrust? In: *The Ashgate Research Companion to New Public Management*. 2010: 309–20. Available: https://www.

academia.edu/26361196/The_Ashgate_Research_Companion_to_New_Public_Management_by_Tom_Christensen_and_Per_Laegreid
4 Peters DH, Noor AA, Singh LP, *et al*. A balanced scorecard for health services in Afghanistan. *Bull World Health Organ* 2007;85:146–51.
5 Edward A, Kumar B, Kakar F, *et al*. Configuring balanced scorecards for measuring health system performance: evidence from 5 years' evaluation in Afghanistan. *PLoS Med* 2011;8:e1001066.
6 Andersen CT, Ahmadzai H, Rasekh AW, *et al*. Improving health service delivery in conflict-affected settings: lessons from a nationwide strategic purchasing mechanism in Afghanistan. *J Glob Health* 2021;11:04049.
7 Alba S, Sondorp E, Kleipool E, *et al*. Estimating maternal mortality: what have we learned from 16 years of surveys in Afghanistan *BMJ Glob Health* 2020;5:e002126.
8 Malik MA, Van de Poel E, Van Doorslaer E. Did contracting effect the use of primary health care units in Pakistan *Health Policy Plan* 2017;32:1032–41.
9 Liu X, Hotchkiss DR, Bose S. The effectiveness of contracting-out primary health care services in developing countries: a review of the evidence. *Health Policy Plan* 2008;23:1–13.
10 Batley R, Mcloughlin C. Engagement with non-state service providers in fragile States: reconciling state-building and service delivery. *Dev Policy Rev* 2010;28:131–54.
11 Soeters R, Griffiths F. Improving government health services through contract management: a case from Cambodia. *Health Policy Plan* 2003;18:74–83.
12 Danforth K, Ahmad AM, Blanchet K, *et al*. Monitoring and evaluating the implementation of essential packages of health services. *BMJ Glob Health* 2023;8:e010726.
13 Palmer N, Strong L, Wali A, *et al*. Contracting out health services in fragile States. *BMJ* 2006;332:718–21.
14 Bertone MP, Jowett M, Dale E, *et al*. Health financing in fragile and conflict-affected settings: what do we know, seven years on? *Soc Sci Med* 2019;232:209–19.
15 Eyben R. *International Aid and the Making of a Better World: Rosalind Eyben*. London and New York: Routledge Taylor and Francis Group, 2014.
16 Finlay L. Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. *Qualitative Research* 2002;2:209–30.
17 Ozano K, Khatri R. Reflexivity, positionality and power in cross-cultural participatory action research with research assistants in rural Cambodia. *Educational Action Research* 2018;26:190–204.
18 World Bank Group. *Afghanistan - System Enhancement for Health Action in Transition Project (English)*. Washington, D.C, 2019. Available: http://documents.worldbank.org/curated/en/841621549319037855/Afghanistan-System-Enhancement-for-Health-Action-in-Transition-Project
19 World Bank Group. Afghanistan - Sehatmandi project (English). 2018. Available: http://documents.worldbank.org/curated/en/619701522461635967/Afghanistan-Sehatmandi-Project
20 Alba S, Straetemans M. Whatever can go wrong, need not go wrong: open quality approach for epidemiology. *Emerg Themes Epidemiol* 2021;18:8.
21 Bertone MP, Jacobs E, Toonen J, *et al*. Performance-based financing in three humanitarian settings: principles and pragmatism. *Confl Health* 2018;12:28.
22 Sagmeister E, Steets J. The use of third- party monitoring in insecure contexts. Secure access in volatile environments (SAVE). 2016. Available: https://assets.publishing.service.gov.uk/media/5f82d1cf8fa8f5044f333664/SAVE__2016__The_use_of_third-party_monitoring_in_insecure_contexts.pdf
23 Newbrander W, Ickx P, Feroz F, *et al*. Afghanistan's basic package of health services: its development and effects on rebuilding the health system. *Glob Public Health* 2014;9 Suppl 1:S6–28.
24 World Health Organization. Regional office for the Eastern Mediterranean. Funding pause results in imminent closure of more than 2000 health facilities in Afghanistan. n.d. Available: http://www.emro.who.int/afg/afghanistan-news/funding-pause-results-in-shut-down-of-more-than-2000-health-facilities-in-afghanistan.html
25 Arur A, Peters D, Hansen P, *et al*. Contracting for health and curative care use in Afghanistan between 2004 and 2005. *Health Policy Plan* 2010;25:135–44.
26 Loevinsohn B, Sayed GD. Lessons from the health sector in Afghanistan: how progress can be made in challenging circumstances. *JAMA* 2008;300:724–6.
27 Roberts B, Guy S, Sondorp E, *et al*. A basic package of health services for post-conflict countries: implications for sexual and reproductive health services. *Reprod Health Matters* 2008;16:57–64.

Supplementary file 1: Round table and interview discussion guide

OECD criteria	MOPH	SPs	TPM
Relevance: Is there a need?	In your view, what was the rationale for having a TPM as an independent party assessing the performance of health service providers? Do you agree with the professed need for this? Moving forward (post 2021), based on your knowledge and experience, do you think alternative models could have been better able to provide information regarding performance and progress of health service providers? Why? How so? What would be their advantages and disadvantages compared to the TPM model?	In your view, what was the rationale for having a TPM as an independent party assessing the performance of health service providers? Do you agree with the professed need for this? Moving forward (post 2021), based on your knowledge and experience, do you think alternative models could have been better able to provide information regarding performance and progress of health service providers? Why? How so? What would be their advantages and disadvantages compared to the TPM model?	In your view, what was the rationale for having a TPM as an independent party assessing the performance of health service providers? Do you agree with the professed need for this? Moving forward (post 2021), based on your knowledge and experience, do you think alternative models could have been better able to provide information regarding performance and progress of health service providers? Why? How so? What would be their advantages and disadvantages compared to the TPM model?
Efficiency: were resources used wisely?	Were the resources (financial, HR, time) put in place for TPM justified compared to the costs of service delivery? Did TPM activities take up much of your time? If yes, could you give an example? Was this useful or less so? How? What are/were the consequences of this? Were there any unintended consequences in terms of (in)efficiency resulting from the existence of a TPM?	Did TPM data/reports support or hinder BPHS/EPHS implementation? How so? Did TPM activities take up much of your time? If yes, could you give an example? Was this useful or less so? How? What are/were the consequences of this? Were there any unintended consequences in terms of (in)efficiency resulting from the existence of a TPM? If considered a hindrance, how could the TPM have functioned in a more supportive manner?	What opportunities do you see for efficiency gain if any, in terms of TPM data collection, data analyses, data validation and report writing?
Effectiveness: was the intervention delivered as planned?	(How) did TPM data/reports allow the identification of strengths and weaknesses in the health system, with regards to the BPHS and EPHS implementation? Do you think TPM data was trustworthy and reliable? Why (examples)? How did that affect your behaviour?	How did TPM data/reports allow identification of strengths and weaknesses in the health system, with regards to the BPHS and EPHS implementation? Do you think TPM data was trustworthy and reliable? Why (examples)? How did that affect your behaviour?	How did TPM data/reports allow the identification of strengths and weaknesses in the health system, with regards to the BPHS and EPHS implementation? Do you think TPM data was trustworthy and reliable? Why (examples)? How did that affect your behaviour?

Alba S, *et al. BMJ Glob Health* 2023; 8:e013470. doi: 10.1136/bmjgh-2023-013470

OECD criteria	MOPH	SPs	TPM
Impact: did it change the lives of the target group?	Was TPM data used to improve BPHS and EPHS implementation? Data from which products/aspects of TPM (HMIS verification, BSC, DQA, AHS)? How? If not, why? What strategies could have helped improve this? Did the existence and practices of a TPM have any (positive or negative) unintended consequences for service delivery and access to (quality) healthcare? What would have happened without any TPM?	Was TPM data used to improve BPHS and EPHS implementation? Data from which products/aspects of TPM (HMIS verification, BSC, DQA, AHS)? How? If not, why? What strategies could have helped improve this? Did the existence of a TPM have any (positive or negative) unintended consequences for service delivery and access to (quality) healthcare? What would have happened without any TPM?	Was TPM data used to improve BPHS and EPHS implementation? How? If not, why? What strategies could have helped improve this? What would have happened without any TPM?
Sustainability: what remains after leaving?	What strategies, if any, were put in place to ensure the systems put in place by TPM would be used in the future by MOPH or SPs? (How) has the existence of a TPM contributed to systems-strengthening? Which factors played an important role in this? Did TPM practices contribute to or undermine aid effectiveness: country ownership, alignment and harmonisation, transparency and accountability? Which factors played an important role in this? Would alternative models have equally or better contributed to health system-strengthening and aid effectiveness?	What strategies, if any, were put in place to ensure the systems put in place by TPM would be used in the future by MOPH or SPs? (How) has the existence of a TPM contributed to systems-strengthening? Which factors played an important role in this? Did TPM practices contribute to or undermine aid effectiveness: country ownership, alignment and harmonisation, transparency and accountability? Which factors played an important role in this? Would alternative models have equally or better contributed to health system-strengthening and aid effectiveness?	What strategies, if any, were put in place to ensure the systems put in place by TPM would be used in the future by MOPH or SPs? (How) has the existence of a TPM contributed to systems-strengthening? Which factors (played an important role in this? Did TPM practices contribute to or undermine aid effectiveness: country ownership, alignment and harmonisation, transparency and accountability? Which factors played an important role in this? Would alternative models have equally or better contributed to health system-strengthening and aid effectiveness?