

TAJIKISTAN FASTR SURVEY: IDENTIFICATION RISK ASSESSMENT AND DATA ANONYMIZATION

1. Introduction

This anonymization plan outlines the procedures to protect the confidentiality of facilities and respondents in the Tajikistan FASTR Health Facility Phone Survey. While the survey data does not directly identify individual patients, it includes several variables that may allow indirect identification of health facilities or personnel. To mitigate this risk, this plan identifies sensitive variables and defines methods for anonymization prior to any data sharing or publication.

2. Overview of Tajikistan FASTR survey

The FASTR survey is a rapid phone-based assessment conducted with 598 primary healthcare facilities in Tajikistan. It gathers data on service delivery and readiness, covering areas such as basic infrastructure, human resources, service delivery, facility financing, medical supplies and equipment, community engagement, leadership and coordination, and quality improvement. The survey was conducted on a nationally representative sample of Family Medicine Departments of City Health Centers (CHCs) and District Health Centers (DHCs), rural health centers (RHCs), and health houses.

3. Key variables with disclosure risk

In line with Statistical Disclosure Control (SDC) protocol¹, the following variables have been identified as posing high or moderate risk for facility or respondent re-identification. These key variables could potentially be used by unauthorized individuals to re-identify facilities. Table 1 presents these variables along with their associated risk levels and justifications.

Table 1: Key variables with disclosure risk in the FASTR survey

Variable	QID / Source	Risk Level	Justification
Respondent name	ID_RESP_NAME	High	Direct identifier
Health facility code	ID_FAC	High	Direct identifier, unique ID

¹ Templ, Matthias, Bernhard Meindl, Alexander Kowarik, and Shuang Chen. "Introduction to Statistical Disclosure Control (SDC)." IHSN Working Paper No. 007 (2014).

Variable	QID / Source	Risk Level	Justification
Health facility name	ID_FAC_NAME	High	Direct identifier
Region name	ID_ADMIN1_NAME	Moderate	Combined with facility type and services may lead to re-identification
District name	ID_ADMIN2_NAME	Moderate	Small districts with few facilities increase re-identification risk
Facility type	ID_FAC_TYPE	Moderate	May identify facilities when combined with administrative levels
Free-text responses	e.g., SERV_04B_TXT, *_OTHER	High	May contain names or unique descriptors
Additional financing/human resources administrative data merged with survey database	HR_*_ADM, FIN_*_ADM	Moderate	Specific staffing or financial details may enable identification

4. Anonymization and disclosure control methods

Anonymization procedures will, first and foremost, aim to ensure that individual health facilities and respondents are not directly identifiable. This is in line with international standards of data sharing and publication, the ethical/IRB approval that has been granted for this study, and the conditions outlined in the informed consent process (see Appendix A).

Category	Method	Variables Covered	Details
Direct identifiers	Remove entirely	ID_RESP_NAME, ID_FAC_NAME	These will be removed from all shared datasets

Category	Method	Variables Covered	Details
	Anonymize	ID_FAC	Replace with pseudonymized IDs (e.g., FAC001, FAC002).
Administrative units	Keep/no change	ID_ADMIN1_NAME	This will be kept without change
	Remove entirely	ID_ADMIN2_NAME	This will be removed from all shared datasets
Facility type	Collapse categories	ID_FAC_TYPE	DHC and CHC combined
Text fields	Redact / tokenize	e.g., SERV_04B_TXT, *_OTHER	Tokenize with general categories
Additional datasets merged in	Remove entirely	HR*_ADM, FIN*_ADM	These will be removed from all shared datasets

Commented [VA1]: Note that this has already been done by the survey firm

Appendix A: Informed consent script

Enumerator reads to respondent: Hello. My name is [interviewer name]. I am calling on behalf of M-Vector, a research company, in coordination with the Ministry of Health and the support of the World Bank and as in accordance with Order № 212 from 12.04.2024. M-Vector is conducting a facility assessment to assist the government in knowing more about the challenges this facility faces in providing health services to your community. This facility was randomly selected to participate in this study. The information collected about this facility may be used by M-Vector, organizations supporting services in this facility, and researchers for planning service improvement or conducting further studies of health services. **The data and findings from this study may be shared in personally de-identified datasets, reports, and/or publications by the Ministry of Health and/or its partners. The responses you provide will be anonymous, and the name of the facility, your name, or the names of any other staff who participated in this study will not be included in the dataset or any report.**

We are asking for your help in collecting this information. You may refuse to answer any question or choose to stop the interview at any time. However, we hope you will answer the questions, which will benefit the services you provide and the nation. If there are questions for which someone else is the most appropriate person to provide the information, we would appreciate it if you introduced me to that person to help us collect that information. At this point, do you have any questions about the study?