

SHORT COMMUNICATION

COLLECTING & DISSEMINATING COVID-19 DATA - FROM FRUSTRATION TO FRUITION

Nausheen Bakht, Syed Shahid Nafees Zaidi*, Salman Saleem**, Shazia Nisar**, Syeda Aliza Shahid***

Armed Forces Post Graduate Medical Institute/National University of Medical Sciences (NUMS) Rawalpindi Pakistan, *Armed Forces Institute of Cardiology/National University of Medical Sciences (NUMS) Rawalpindi Pakistan, **Pak Emirates Military Hospital/National University of Medical Sciences (NUMS) Rawalpindi Pakistan, ***Combined Military Hospital Lahore Medical College/National University of Medical Sciences (NUMS) Rawalpindi Pakistan

ABSTRACT

In the unprecedented COVID-19 situation, Medical Directorate needed to get near-real-time data for decision making and policy formulation. Crisis Management Cell was hence established with a vision to act as a data collection point and repository for Armed Forces COVID-19 PCR positive cases. Its foremost task was transforming a traditionally passive system for disease surveillance to a near-real-time semi-automated one. This coupled with the ensuing workload in the coming days especially during surge, constituted a formidable challenge. Through this work we want to share our hands-on experience of data collection & dissemination at Crisis Management Cell. We aim to inform our readers about the important points that need to be taken into account when establishing disease surveillance centers. As per our experience, these include an assessment of demand for near-real-time data, data handlers, data architecture, data security, data quality, data record, data velocity, data variety, data variability, data documentation, data reconciliation and reflection.

Keywords: Data collection, Data dissemination, Data quality, Disease surveillance, Real-time data.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

There cannot be two views about curbing disease outbreaks in their very tracks. To achieve this lofty goal, alignment of public health resources, scientific expertise, relevant experience, and political commitment is mandatory¹. An essential component of the public health infrastructure and an essential resource in identifying, comparing, and predicting disease outbreaks is high-quality surveillance data i.e. data that has been transformed into meaningful information. During a pandemic the significance of having valid, reliable, and line listed data² increases manifold not only for assessing disease infection & case-fatality rates³ but also for crafting appropriate public health response. This often requires mobilizing accurate, voluminous, and compatibly formatted data⁴.

Real-time reporting is fast becoming the

norm in organizations that desire to make timely and data-driven decisions with no delay in delivering information. The exemplars include financial institutions, utilities, telecommunication, healthcare, hospitality industry, customer service, etc which define data by volume, velocity, variety, variability, veracity, and complexity for making data-driven and evidence-backed decisions for policy formulation⁵. In healthcare, it is necessary to collect and process data about what happened and the most critical drivers of the disease⁶. However, collecting and disseminating an overwhelming amount of data in an urgent hour-by-hour manner is quite challenging⁷. It is known to induce shortcuts that compromise quality, ethical standards and an 'analysis paralysis'^{8,9}.

Crises are known to make room for opportunities. The current COVID-19 pandemic has been successful in raising the awareness of all and sundry in healthcare (and elsewhere) about the importance of data in decision making, capacity development in accessing and using

Correspondence: Dr Nausheen Bakht, Assistant Professor, Armed Forces Post Graduate Medical Institute, Rawalpindi Pakistan
Received: 21 Aug 2020; revised received: 28 Aug 2020; accepted: 31 Aug 2020

Table: Challenges & lessons learned in COVID-19 data collection & dissemination at crisis management cell (CMC).

Points to Consider	Lessons Learned What Worked / What did not / What could
Demand for Near-Real-Time Data	<p>Ours is a traditional passive reporting system. In these unprecedented times, stakeholders' expectations for near-real time data were not misplaced but constituted a tall order.</p> <p>We recognized early that technology will be our ally in the fight against COVID-19 and attempted to address our stakeholders' needs by adopting the fastest possible means like phone, Fax, SMS and twice-daily reporting. Since reporting via the Internet was not an option for security reasons therefore formats like Whatsapp and email were ruled out. To date much progress has been made in bringing a real-time disease dashboard in collaboration with C4I to fruition.</p>
Data System Design	<p>Every data has a life cycle¹² comprising of following iterative stages (spanning from its generation to final archival and/or deletion):</p> <p>Generation: Data is acquired into the organization (usually through data entry).</p> <p>Maintenance: Data is processed prior to use.</p> <p>Use: Data is actively used to support the organization's aims and objectives.</p> <p>Publication: Data is sent outside the organization.</p> <p>Archiving: Data is stored for future reference.</p> <p>Purging: Once it is archived, all copies of data may be deleted.</p> <p>We designed our system duly keeping in view all stages of data life cycle. We were duly cognizant of quality assurance and quality control processes as we aimed to prevent errors from entering and/ or staying in our data set through:</p> <ul style="list-style-type: none"> • Data Quality Assurance (all actions carried out before data collection) • Data Quality Control (all actions carried out during & after data collection) <p>Before</p> <ul style="list-style-type: none"> • We developed consensus to clearly define data standards, measurement units, formats, abbreviations and metadata up front. • We duly assigned data associated responsibilities to our trained personnel enabling them to amicably sort data fields and check for discrepancies and outliers at a glance. <p>During</p> <ul style="list-style-type: none"> • We restricted what could be entered into the database by carefully crafting response options for most cells in our spreadsheet. • We deliberated on using consistent terminology and abbreviations. • We ensured that there would only be one piece of information per cell to help eliminate mistakes during data entry. • We introduced color coding for deaths, on vent, recovered and new cases for ease of our data recipients in order to document changes in data. <p>After</p> <ul style="list-style-type: none"> • We passed instructions to our trained personnel to daily eyeball all MS-Excel sheets received from our downstream data recipients in order to quickly sort data fields by checking for discrepancies i.e. any missing, impossible, or anomalous values, if any.
Data Handlers	<p>CMC team was heterogeneously constituted. Only few of its members had a public health background.</p> <ul style="list-style-type: none"> • We remained cognizant of the fact that not all our members had a Public Health background. We therefore invested in identifying and addressing the training needs of our workforce duly taking into account the learning curve of new members. • We kept the composition of our team balanced when we formulated the duty rosters. • We deputed a heterogeneous team of experienced and novice personnel having Public Health and non-Public Health background to compile Morning & Evening COVID-19 Situation Reports (SitReps).
Data Architecture	<p>It stands for SOPs which inform what data will be collected, how it will be stored, integrated and utilized. In the past six months, our data architecture transformed several times especially in the days before the COVID-19 surge. The drivers for these modifications were either higher authorities' requirements or our own experience.</p> <p>We were able to establish well-defined data architecture through dialog. Although this mutual agreement with an overwhelming majority of our stake holders took some time to establish but once settled, it largely remained unaltered in the upcoming days.</p>

Data Security	The problems of compromised data security, impingement of data privacy and the likelihood of confidentiality breaches need to be deliberated upon thoroughly.
	<ul style="list-style-type: none"> • We designed an MS-Excel Sheet which was OAS based. • We kept our soft folders password protected and our hard copies under lock and key and this duty was assigned to a dedicated Officer and a Clerk. • We mentioned the names of team members who formulated, verified and approved the SitReps in writing to apportion accountability accordingly.
Data Quality	Incorrect data entries are said to result in a 'domino effect' as wrong information spreads and does not remain limited to one separate spreadsheet. Moreover 'data contamination' also results in erroneous values in the data set. Although our stakeholders did an excellent job by providing complete, accurate, and timely data, however, at times duplications/redundancies/inconsistencies in data would creep up.
	<ul style="list-style-type: none"> • We overcame the issue of data quality by disseminating detailed instructions to all downstream stakeholders regarding the compilation of MS-Excel sheet for data collection. • We conducted the daily audit of data for discovering these faults. We then explained all inconsistencies by adding relevant explanatory notes in daily SitReps, where warranted.
Data Record	CMC's initial database was MS-Word based which offered limited help for a short while during the pandemic. The files become heavier and took considerably long time in opening up.
	<ul style="list-style-type: none"> • We were agile in switching to the MS-Excel which was shared with all our downstream stakeholders. • We designed our customized MS-Excel Sheet to have a standard format with carefully crafted built-in drop box menus. • We compiled detailed instructions for filling in the MS-Excel Sheet and shared them with our end users.
Data Velocity	The twice-daily COVID-19 SitReps were formulated by CMC on the basis of data received from Armed Forces healthcare facilities across the country. The sentinel events were required to be reported by the focal persons deputed by the facilities within 2 hours.
	We were able to clearly define what information our upstream stakeholders' desire by dialog and directive. Feedback to our downstream data recipients however remained a weak link largely due to the rapid turn-over of data.
Data Variety	2 x SitReps and a One-Pager Summary constituted our daily work products in addition to a Weekly Analysis Report.
	We also compiled data on confirmed, suspected and quarantined individuals. Cumulative and new counts were noted on facility, Corps and Arms wise basis. CMC members also made an effort to contribute towards writing research articles on COVID-19.
Data Variability	Our true challenge was remaining on top of on-vent, discharged or recovered individuals' data. Deaths being sentinel events did not pose a pressing issue.
	We divided Corps Statistics and our several databases (health care workers, deceased individuals, on vent cases, persons on novel therapies) among the team members for daily updating.
Data Documentation	The deluge and velocity of data and the frequency of reporting often mounted to frenzy especially during the surge which forbade keeping detailed documentation of changes in procedures.
	We regularly reflected upon what went well and what did not go as planned. The team consulted among themselves and developed consensus for course correction where warranted. Although detailed documentation was not possible we did formulate customized checklists for our better functioning.
Data Reconciliation	All concerned made their utmost efforts to safeguard the veracity of data. Although few and far in between, due to the paucity of time and in the face of a rapidly evolving situation, this was often not timely possible.
	We issued disclaimers and elaborate explanations for all concerned whenever such dire measures were resorted to.
Reflection	Due to the quickly evolving situation, reports generated very thoughtfully at one point in time, lost their utility rapidly. The additions and deletions in them resulted in a change of format. It was also pointless to obtain same information through multiple reports & formats. Such an arrangement was clearly inefficient but increased the workload unnecessarily. However, some facilities continued sending obsolete formats which made data standardization cumbersome.
	We regularly deliberated upon our existing reports and returns and did course correction by cancelling the ones no longer required.

information, and raising the demand for evidence. It is indeed an excellent opportunity for digital technology¹⁰.

Transforming any disease data collection & dissemination system is an uphill task. Crisis Management Cell (CMC) was established with a vision to act as a data collection point and repository for COVID-19 cases by Medical Directorate. This involved a paradigm shift from an existing traditional/passive reporting system to a near-real-time semi-automated arrangement. Its complexities increased in the wake of pandemic's fast evolution which warranted an at par revision of SOPs in sync with the emerging evidence¹¹. From the point of view of data collection & dissemination, CMC encountered much frustration and successfully converted it into fruition. We are writing this article to share our first-hand experience of collecting and disseminating COVID-19 data in the Armed Forces. Best practices need to be implemented before, during and after data collection to ensure high-quality data. Our reflection reveals what worked, what did not, and what could. Challenges faced and lessons learned in COVID-19 data collection & dissemination at CMC are as per table.

CONCLUSION

We opine that if there is a silver lining associated with COVID-19, it is the successful sensitization of all healthcare establishments and others about the importance of disease data collection & dissemination especially in real-time. This demands that all concerned act in the spirit of 'carpe diem' and endeavor to create a culture

where everyone is a data steward and assumes responsibility for data accuracy, completeness, consistency and integrity.

CONFLICT OF INTEREST

This study has no conflict of interest to be declared by any author.

REFERENCES

1. Kapata N, Ihekweazu C, Ntoumi F, Raji T, Chanda-Kapata P, Mwaba P, et al. Is Africa prepared for tackling the COVID-19 (SARS-CoV-2) epidemic. Lessons from past outbreaks, ongoing pan-African public health efforts, and implications for the future. *Int J Infect Dis* 2020; 93(1): 233-36.
2. Xu B, Kraemer MUG. Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect Dis* 2020; 20(5): 534-39.
3. Morabia A. The Public Health We Need. *Am J Public Health* 2020; 110(7): 923-24.
4. Bernier A, Knoppers BM. Pandemics, privacy, and public health research. *Can J Public Health* 2020; 111(4): 488-58.
5. Agrebi S, Larbi A. Use of artificial intelligence in infectious diseases. In: *Artificial Intelligence in Precision Health*. Elsevier 2020: 415-38. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9780128171332000185>
6. Nardini S, Sanguinetti CM, De Benedetto F, Baccarani C, Del Donno M, Polverino M, et al. SARS-CoV-2 pandemic in Italy: ethical and organizational considerations. *Multidiscip Respir Med* 2020; 15(1): 672-78.
7. Banerjee DA, Mansfield A. Shining a light on the gaps for learning. *Midwifery* 2020; 89: 102790.
8. Coronavirus COVID-19 (2019-nCoV). 2020. Available from: <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>
9. Moore JH, Barnett I, Boland MR, Chen Y, Demiris G, Gonzalez-Hernandez G et al. Ideas for how informaticians can get involved with COVID-19 research. *Bio Data Min* 2020; 13(1): 3-11.
10. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. *Nat Med* 2020; 26(4): 459-61.
11. Hechenbleikner EM, Samarov DV, Lin E. Data explosion during COVID-19: A call for collaboration with the tech industry & data scrutiny. *E Clin Med* 2020; 23(1): 100377.
12. Data life cycle. Whatls.com. 2020. Available from: <https://whatis.techtarget.com/definition/data-life-cycle#:~:text=The%20data%20life%20cycle%20is,in%20the%20data%20life%20cycle>