Key Messages Regarding COVID-19 Data Quality

- **The COVID-19 immunization campaign is larger than any vaccine campaign in history, and it requires timely capture and reporting of an unprecedented amount of data from a significantly greater number of administering providers.**
- **Data in immunization information systems (IIS) is consolidated from hundreds of data sources. Data quality can be hampered by incomplete or inaccurate data entry processes at the source (e.g., race is required if known, but many clinics/pharmacies don't record or message it consistently). These issues can take time and collaboration with partners to resolve.**
- **Partners at all levels (e.g., health clinics; hospitals; pharmacies; state, local, and federal public health) are working diligently to improve data quality and enhance data exchange between programs and systems.**

Purpose of This Document

Given the unprecedented nature of the COVID-19 vaccine roll out, there is a need to broaden understanding of data issues and discrepancies when comparing administered vaccination data across multiple systems. These talking points are provided as background for IIS in discussing data quality with their state/jurisdiction leadership, their data exchange partners, and their CDC partners. These talking points represent an effort to better manage expectations of data quality and comparability across systems. It is important to note that this is not an effort to excuse or diminish the importance of data quality challenges, but rather to provide helpful context for interpreting data that may be available in multiple varied systems.

General Messages

- Data discrepancies across various systems are not unexpected. They could be due to multiple factors, many of which are addressed below.
- No one entity has responsibility for the quality of COVID-19 data. This is a shared responsibility among thousands of providers, systems, and levels of government.
- COVID-19 is a novel vaccine with unique characteristics, and it is being rolled out as part of a vaccination campaign of unprecedented size and scope. Given that vaccination just began in December 2020, the routinization of administration and reporting are still ramping up. Increased volume of administrations and reporting may affect data quality.
- When analyzing, interpreting, or otherwise using COVID-19 data, it is beneficial to have input or review from those closest to the data to ensure conclusions are sound. Misinterpretations/assumptions can lead to confusing messaging.

Completeness:

- The reporting requirements within the COVID-19 response are complex and multiple systems (electronic health record systems, pharmacy systems, IIS, CDC Data Clearinghouse, CDC Data Lake, Tiberius, COVID Data Tracker, etc.) are involved and engaged in collecting data on doses administered and reported. Some administering entities, such as pharmacies, are dual

reporters, reporting doses to both an IIS and the CDC Data Clearinghouse (DCH), which, if not accounted for, could result in differing information in state and federal systems.

- Some IIS may have queues or backlogs of data exchange partners waiting to onboard. These sites may be directed to enter doses manually through an IIS web-based user interface while they await the development of an automated interface, but this manual step may adversely affect complete data capture in the short term.
- Many data fields in IIS are standardized across jurisdictions, but not all fields are mandatory, such as race and ethnicity. Race and ethnicity are required if known, and may not be thoroughly captured as part of the workflow in pharmacies, clinics, etc. In addition, "unknown" is an allowable entry, so even if the field is populated, it may not include meaningful data sufficient to answer health equity questions. Some jurisdictions have state and local policies related to data privacy that prohibit the capture or exchange related to an individual's race or ethnicity.

Accuracy:

- Various systems may have differing inclusion/exclusion criteria. For example, an IIS may receive reports of doses administered in another jurisdiction, but these doses may be excluded from the redacted data they report to CDC. This can create discrepancies in overall numbers, but is an expected occurrence based on inclusion/exclusion criteria.
- Data quality interventions may occur within various systems, resulting in corrections or modifications. This can result in differing data across systems.
- Some systems don't currently accept/allow for corrections, which means some reporting errors may be corrected upstream, but downstream systems don't have a way to accept those corrections.

Timeliness:

- Per the CDC website, "COVID-19 vaccination providers must document vaccine administration in their medical record systems within 24 hours of administration, and use their best efforts to report administration data to the relevant system for the jurisdiction (i.e., IIS) as soon as practicable and no later than 72 hours after administration." Once the data reaches the IIS, they submit it to the CDC DCH by noon local time the following day. Data then moves from the CDC Data Clearinghouse, to the CDC Data Lake, to Tiberius, and to the CDC COVID Data Tracker on an internal schedule. Some differences in data should be anticipated as data may be at various points in this system.

- Data quality interventions can take additional time but may result in better data being reported. For example, if an IIS receives a message that includes an illogical vaccine code (CVX/MVX), they may need time to follow up with a provider facility to correct the error. It is to everyone's benefit if this correction happens before the data are reported to another system in the reporting chain. Similarly, the IIS may receive an appropriate vaccine code (CVX/NDC) in a message but no manufacturer (MVX). The IIS may take additional time to infer the manufacturer (MVX) from the vaccine code (CVX/NDC), thus improving the overall data quality of the message before it is sent on.