

1. Discuss in detail the likely causes of data quality issues in systems used by the following (5 points):

a.) primary care provider

Data quality issues in primary care settings often stem from a combination of human, technical, and organizational factors. Many clinics still rely heavily on manual data entry, but documentation can be a source of frustration and workflow disruption. Poor usability, such as unintuitive navigation and lack of autosave, adds cognitive load and reduces system transparency (Cahill et al., 2025; Patel et al., 2021). These issues lead to inefficiency and increase the likelihood of documentation errors like incomplete, inconsistent, or incorrect entries (Cahill et al., 2025; Patel et al., 2021). This is especially exacerbated with paper records, which 22.8% of office-based physicians in the U.S. still use (National Center for Health Statistics, 2024), raising the risk of duplicate, missing, or fragmented patient records (Cruz-Correia et al., 2009).

The absence of automated validation checks (e.g., enforcing proper date or unit formats) further degrades data consistency (CCD Health, 2025; Madandola et al., 2023). Data standards can also vary between electronic health record (EHR) systems, particularly if they are legacy or outdated, preventing seamless data exchange between labs, pharmacies, hospitals, and other external systems (Bakken et al., 2021). Moreover, about 10.4% of office-based physicians still use uncertified EHRs (National Center for Health Statistics, 2024), which may further hinder compatibility and data-sharing capabilities (Bakken et al., 2021). These issues are more common in small or rural practices, where the high financial investment required for EHR implementation often delays system updates and may drive the adoption of uncertified, lower-cost alternatives (Felt-Lisk et al., 2010). Lastly, insufficient staff training and ongoing support limit the ability to identify and correct errors in real time, reducing both the accuracy and timeliness of patient data and ultimately impacting care quality (Bakken et al., 2021; CCD Health, 2025).

b.) multi-country NGO

For multi-country non-governmental organizations (NGOs), data quality challenges often arise from cross-cultural, infrastructural, and contextual complexities (Gharawi et al., 2009). Language barriers can cause translation inaccuracies or semantic loss, especially when unstructured data (e.g., field notes, interviews) must be standardized across regions (Gazi, 2020; Gharawi et al., 2009). Different countries may use inconsistent data definitions or classification schemes, complicating data integration and comparison (Gharawi et al., 2009).

Technical issues also play a major role. NGOs working across low-resource settings may face unstable internet connectivity or limited access to modern data collection tools, potentially impacting data accuracy, volume, and timeliness of sharing (Gharawi et al., 2009, van Zyl et al., 2021; Whalen et al., 2014). Additionally, NGOs operating in high-stress or crisis environments (e.g., disaster relief zones) often collect data under urgent conditions with limited oversight, which increases the risk of inaccurate or incomplete field reporting (Gazi, 2020; Masefield et al., 2020). Strengthening data governance,

Group 3

B585 – Biomedical Analytics - Assignment 3

standardizing terms and classifications, and investing in system interoperability would help address these root causes and enhance data quality across all contexts.

2. Data Validation Exercise - write the validation rules for all 4 questions below (10 points)

I. Gender Based Violence		Person 1		Person 2		Person 3		Person 4	
		Under 5 years		5-9 years		10-18 years		19 years and above	
		M	F	M	F	M	F	M	F
1	GBV victims with symptoms of sexual violence (new cases).								
2	GBV victims with symptoms of physical violence (new cases).								
3	GBV victims referred for care to higher level health facility								
4	GBV victims referred to this facility by police								
5	GBV victims referred to this facility by community health workers								
6	GBV victims HIV+ sero-conversion 3 months after exposure								
7	GBV victims with irreversible disabilities due to GBV								
8	GBV victim deaths								
9	GBV victims pregnant 4 weeks after exposure								
10	GBV victims received emergency contraception within 72 hours								
11	GBV victims received post exposure HIV prophylaxis within 48 hours								

1. List at least 4 validation rules that could be created between the different data elements – put your list in a document to upload to oncourse (G# TS validation rules)

Assumptions:

1. Sexual violence (Element 1) is a subset of physical violence (Element 2)—as stated in Lecture 8.
2. Data will be aggregated from complete reporting periods, with a minimum three-month lag after initial case reporting to allow time for HIV test results.
3. HIV is assumed to be contracted from sexual violence. Although HIV can be contracted from physical violence (blood to blood contact), it is uncommon enough that testing would only occur for sexual violence.

Group 3

B585 – Biomedical Analytics - Assignment 3

1a. The number of GBV victims that are HIV Positive 3 months after exposure (Element **6**)
≤ the number of GBV victims with symptoms of sexual violence (Element **1**).

Rules:

Left hand side	Operator	Right hand side
GBV victims w/ symptoms of physical violence (<5y, M)	>=	GBV victims HIV+ sero-conversion 3 months after exposure(<5y, M)
GBV victims w/ symptoms of physical violence (<5y, F)	>=	GBV victims HIV+ sero-conversion 3 months after exposure(<5y, F)
GBV victims w/ symptoms of physical violence (5-14y, M)	>=	GBV victims HIV+ sero-conversion 3 months after exposure(5-14y, M)
GBV victims w/ symptoms of physical violence (5-14y, F)	>=	GBV victims HIV+ sero-conversion 3 months after exposure(5-14y, F)
GBV victims w/ symptoms of physical violence (10-18y, M)	>=	GBV victims HIV+ sero-conversion 3 months after exposure(10-18y, M)
GBV victims w/ symptoms of physical violence (10-18y, F)	>=	GBV victims HIV+ sero-conversion 3 months after exposure(10-18y, F)
GBV victims w/ symptoms of physical violence (>19y, M)	>=	GBV victims HIV+ sero-conversion 3 months after exposure(>19y, M)
GBV victims w/ symptoms of physical violence (>19y, F)	>=	GBV victims HIV+ sero-conversion 3 months after exposure(>19y, F)

1b. The number of GBV victims with symptoms of physical violence (Element **2**) \geq The number of GBV victim deaths (Element **8**)

Rules:

Left hand side	Operator	Right hand side
GBV victims w/ symptoms of physical violence (<5y, M)	>=	GBV victims deaths(<5y, M)
GBV victims w/ symptoms of physical violence (<5y, F)	>=	GBV victims deaths(<5y, F)
GBV victims w/ symptoms of physical violence (5-14y, M)	>=	GBV victims deaths(5-14y, M)
GBV victims w/ symptoms of physical violence (5-14y, F)	>=	GBV victims deaths(5-14y, F)
GBV victims w/ symptoms of physical violence (10-18y, M)	>=	GBV victims deaths(10-18y, M)
GBV victims w/ symptoms of physical violence (10-18y, F)	>=	GBV victims deaths(10-18y, F)
GBV victims w/ symptoms of physical violence (>19y, M)	>=	GBV victims deaths(>19y, M)
GBV victims w/ symptoms of physical violence (>19y, F)	>=	GBV victims deaths(>19y, F)

Group 3

B585 – Biomedical Analytics - Assignment 3

1c. The number of GBV victims referred to this facility by police and community health workers (Element 4 + Element 5) \leq the GBV victims with symptoms of physical violence (Element 2).

Rules:

Left hand side	Operator	Right hand side
GBV victims w/ symptoms of physical violence (<5y, M)	\geq	(GBV victims referred to this facility by police(<5y, M) + GBV victims referred to this facility by community health workers(<5y, M))
GBV victims w/ symptoms of physical violence (<5y, F)	\geq	(GBV victims referred to this facility by police(<5y, F) + GBV victims referred to this facility by community health workers(<5y, F))
GBV victims w/ symptoms of physical violence (5-14y, M)	\geq	(GBV victims referred to this facility by police(5-14y, M) + GBV victims referred to this facility by community health workers(5-14y, M))
GBV victims w/ symptoms of physical violence (5-14y, F)	\geq	(GBV victims referred to this facility by police(5-14y, F) + GBV victims referred to this facility by community health workers(5-14y, F))
GBV victims w/ symptoms of physical violence (10-18y, M)	\geq	(GBV victims referred to this facility by police(10-18y, M) + GBV victims referred to this facility by community health workers(10-18y, M))
GBV victims w/ symptoms of physical violence (10-18y, F)	\geq	(GBV victims referred to this facility by police(10-18y, F) + GBV victims referred to this facility by community health workers(10-18y, F))
GBV victims w/ symptoms of physical violence (>19y, M)	\geq	(GBV victims referred to this facility by police(>19y, M) + GBV victims referred to this facility by community health workers(>19y, M))
GBV victims w/ symptoms of physical violence (>19y, F)	\geq	(GBV victims referred to this facility by police(>19y, F) + GBV victims referred to this facility by community health workers(>19y, F))

1d. The number of GBV victims with symptoms of sexual violence (Element 1) \geq the number of GBV victims who received post-exposure HIV prophylaxis within 48 hours (Element 11)

Rules:

Left hand side	Operator	Right hand side
GBV victims w/ symptoms of sexual violence(<5y, M)	\geq	GBV victims received post exposure HIV prophylaxis(<5y, M)
GBV victims w/ symptoms of sexual violence(<5y, F)	\geq	GBV victims received post exposure HIV prophylaxis(<5y, F)
GBV victims w/ symptoms of sexual violence(5-14y, M)	\geq	GBV victims received post exposure HIV prophylaxis(5-14y, M)
GBV victims w/ symptoms of sexual violence(5-14y, F)	\geq	GBV victims received post exposure HIV prophylaxis(5-14y, F)
GBV victims w/ symptoms of sexual violence(10-18y, M)	\geq	GBV victims received post exposure HIV prophylaxis(10-18y, M)
GBV victims w/ symptoms of sexual violence(10-18y, F)	\geq	GBV victims received post exposure HIV prophylaxis(10-18y, F)
GBV victims w/ symptoms of sexual violence(>19y, M)	\geq	GBV victims received post exposure HIV prophylaxis(>19y, M)
GBV victims w/ symptoms of sexual violence(>19y, F)	\geq	GBV victims received post exposure HIV prophylaxis(>19y, F)

Group 3

B585 – Biomedical Analytics - Assignment 3

1e. the number of GBV victims with symptoms of physical violence (Element **2**) \geq the number of GBV victims with irreversible disabilities due to GBV (Element **7**).

Rules:

Left hand side	Operator	Right hand side
GBV victims w/ symptoms of physical violence (<5y, M)	\geq	GBV victims w/ irreversible disability due to GBV(<5y, M)
GBV victims w/ symptoms of physical violence (<5y, F)	\geq	GBV victims w/ irreversible disability due to GBV(<5y, F)
GBV victims w/ symptoms of physical violence (5-14y, M)	\geq	GBV victims w/ irreversible disability due to GBV(5-14y, M)
GBV victims w/ symptoms of physical violence (5-14y, F)	\geq	GBV victims w/ irreversible disability due to GBV(5-14y, F)
GBV victims w/ symptoms of physical violence (10-18y, M)	\geq	GBV victims w/ irreversible disability due to GBV(10-18y, M)
GBV victims w/ symptoms of physical violence (10-18y, F)	\geq	GBV victims w/ irreversible disability due to GBV(10-18y, F)
GBV victims w/ symptoms of physical violence (>19y, M)	\geq	GBV victims w/ irreversible disability due to GBV(>19y, M)
GBV victims w/ symptoms of physical violence (>19y, F)	\geq	GBV victims w/ irreversible disability due to GBV(>19y, F)

Group 3

B585 – Biomedical Analytics - Assignment 3

2. Create validation rules comparing data elements 1, 2 and 3 – name (create the rules with your Group # at the beginning of the name, (e.g. G# GBV <5 years etc....)

Rule Name: G3 GBV Referrals to Higher Care Facility

Rule: $E2 \geq E1$ and $E2 \geq E3$

$E2 \geq E1$

Left hand side	Operator	Right hand side
GBV victims w/ symptoms of physical violence (<5y, M)	\geq	GBV victims w/ symptoms of sexual violence (<5y, M)
GBV victims w/ symptoms of physical violence (<5y, F)	\geq	GBV victims w/ symptoms of sexual violence (<5y, F)
GBV victims w/ symptoms of physical violence (5-14y, M)	\geq	GBV victims w/ symptoms of sexual violence (5-14y, M)
GBV victims w/ symptoms of physical violence (5-14y, F)	\geq	GBV victims w/ symptoms of sexual violence (5-14y, F)
GBV victims w/ symptoms of physical violence (10-18y, M)	\geq	GBV victims w/ symptoms of sexual violence (10-18y, M)
GBV victims w/ symptoms of physical violence (10-18y, F)	\geq	GBV victims w/ symptoms of sexual violence (10-18y, F)
GBV victims w/ symptoms of physical violence (>19y, M)	\geq	GBV victims w/ symptoms of sexual violence (>19y, M)
GBV victims w/ symptoms of physical violence (>19y, F)	\geq	GBV victims w/ symptoms of sexual violence (>19y, F)

$E2 \geq E3$

Left hand side	Operator	Right hand side
GBV victims w/ symptoms of sexual violence (<5y, M)	\geq	GBV victims referred for care to higher level health facility(<5y, M)
GBV victims w/ symptoms of sexual violence (<5y, F)	\geq	GBV victims referred for care to higher level health facility(<5y, F)
GBV victims w/ symptoms of sexual violence (5-14y, M)	\geq	GBV victims referred for care to higher level health facility(5-14y, M)
GBV victims w/ symptoms of sexual violence (5-14y, F)	\geq	GBV victims referred for care to higher level health facility(5-14y, F)
GBV victims w/ symptoms of sexual violence (10-18y, M)	\geq	GBV victims referred for care to higher level health facility(10-18y, M)
GBV victims w/ symptoms of sexual violence (10-18y, F)	\geq	GBV victims referred for care to higher level health facility(10-18y, F)
GBV victims w/ symptoms of sexual violence (>19y, M)	\geq	GBV victims referred for care to higher level health facility(>19y, M)
GBV victims w/ symptoms of sexual violence (>19y, F)	\geq	GBV victims referred for care to higher level health facility(>19y, F)

The number of GBV victims with symptoms of physical violence (Element **2**) \geq the number of GBV victims with symptoms of sexual violence (Element **1**). AND The number of GBV victims with symptoms of physical violence (Element **2**) \geq the number of GBV victims referred for care to higher level facility (Element **3**).

3. Create validation rules that compare data elements 1 and 10 (e.g., G# GBV emergency PEP)

Rule Name: G3 GBV Emergency PEP

Rules:

Left hand side	Operator	Right hand side
GBV victims w/ symptoms of physical violence (<5y, M)	>	GBV victims received emergency contraceptive(<5y, M)
GBV victims w/ symptoms of physical violence (<5y, F)	>	GBV victims received emergency contraceptive(<5y, F)
GBV victims w/ symptoms of physical violence (5-14y, M)	>	GBV victims received emergency contraceptive(5-14y, M)
GBV victims w/ symptoms of physical violence (5-14y, F)	>	GBV victims received emergency contraceptive(5-14y, F)
GBV victims w/ symptoms of physical violence (10-18y, M)	>	GBV victims received emergency contraceptive(10-18y, M)
GBV victims w/ symptoms of physical violence (10-18y, FM)	>	GBV victims received emergency contraceptive(10-18y, F)
GBV victims w/ symptoms of physical violence (>19y, M)	>	GBV victims received emergency contraceptive(>19y, M)
GBV victims w/ symptoms of physical violence (>19y, F)	>	GBV victims received emergency contraceptive(>19y, F)

The number of GBV victims who are female and over five years old with symptoms of sexual violence (Element **1**) \geq the number of GBV victims who are female and over five years old receiving emergency contraception (Element **10**).

Group 3

B585 – Biomedical Analytics - Assignment 3

4. Create a validation rule that sends a message to a user group if a GBV victim death is reported (e.g., G# GBV <5 years deaths....)

Rule Name: G3_GBV_DeathsAlertByAge

Rules:

Left hand side	Operator	Right hand side
GBV victim death	>	0
GBV victim death	>	0
GBV victim death	>	0
GBV victim death	>	0
GBV victim death	>	0
GBV victim death	>	0
GBV victim death	>	0
GBV victim death	>	0
GBV victim death	>	0

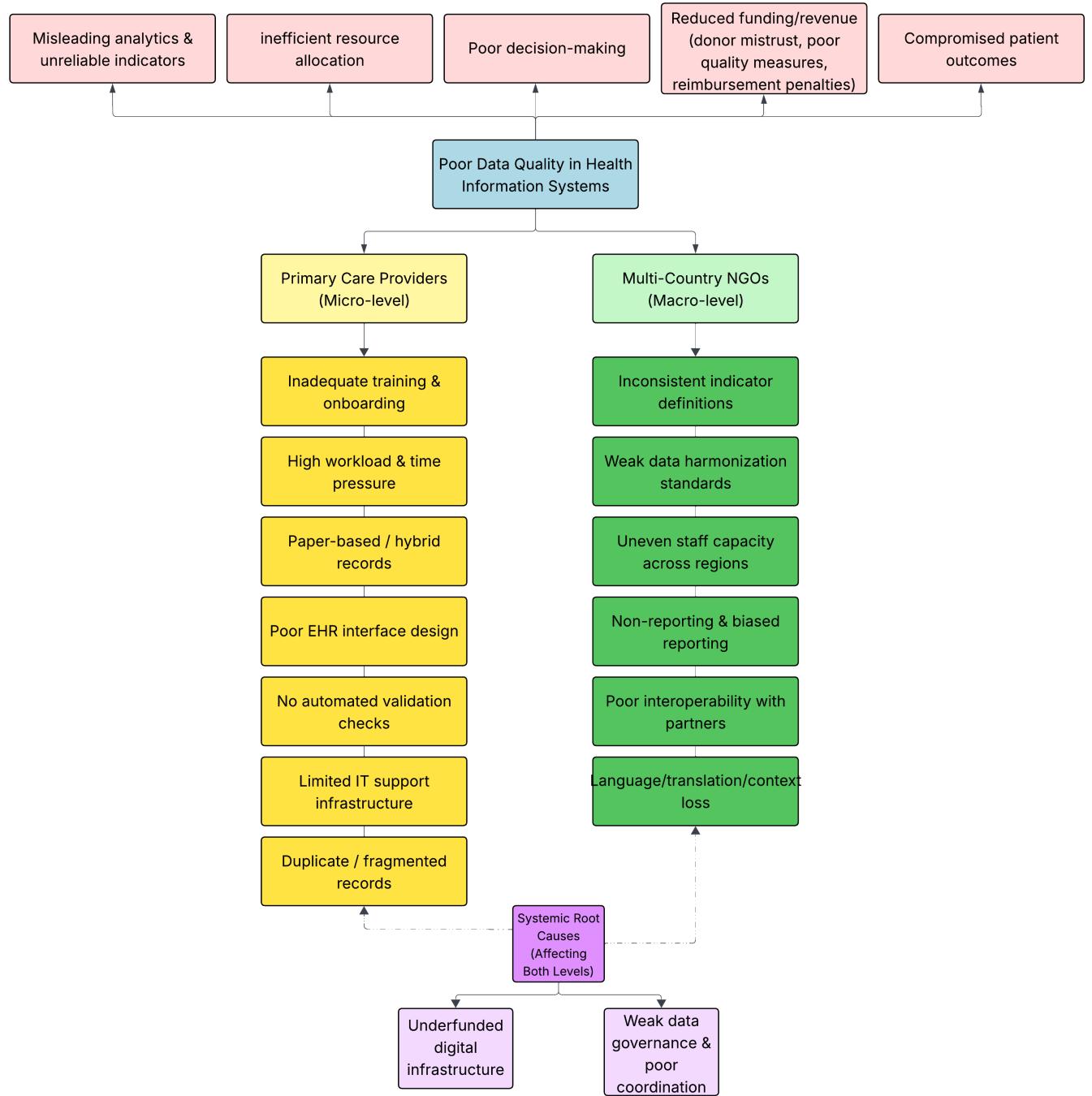
Notification Template:

- **Name:** ALERT: GBV death reported
- **Notification recipient:** GBV_Death_Usergroup
- **Message subject/body:**
 - **Subject:** ALERT: GBV death reported
 - **Body:** A GBV-related death has been reported. Immediate follow-up required by the GBV Monitoring Team.

Group 3

B585 – Biomedical Analytics - Assignment 3

Extra credit (2 points): Within your groups, discuss the ideas that you would like to implement to improve data quality in your organizations. Prepare a problem tree (problem, root causes, consequences) for your group discussions.



Discussed Solutions for the problems we identified

Micro-level

Identified Problem	Proposed Solution / Intervention
Inadequate training & onboarding	Develop continuous training programs and refresher workshops on digital data entry, validation, and coding standards. Include competency-based assessments.
High workload & time pressure	Introduce data-entry support staff or automated data capture (e.g., barcode scanning, speech-to-text). Streamline workflows to minimize redundant reporting.
Paper-based / hybrid records	Gradually transition to fully electronic systems; pilot digital record systems in selected facilities before national rollout.
Poor EHR interface design	Involve end users (clinicians, nurses, data clerks) in interface design; conduct usability testing to ensure intuitive layout and reduced entry errors.
No automated validation checks	Implement built-in data validation rules in the EHR (e.g., date range checks, mandatory fields, logical comparisons). Integrate simple pop-up error alerts.
Limited IT support infrastructure	Establish regional IT helpdesks or on-call technical support. Provide on-site technicians for large health centers.
Duplicate / fragmented records	Use unique patient identifiers (UPIs) or national ID-based linking. Enforce record deduplication algorithms periodically.

Macro-level

Identified Problem	Proposed Solution / Intervention
Inconsistent indicator definitions	Develop a unified metadata dictionary and standard indicator catalog agreed upon by all partners.

Group 3

B585 – Biomedical Analytics - Assignment 3

Weak data harmonization standards	Adopt international data exchange standards (e.g., HL7 FHIR, DHIS2 metadata packages). Conduct joint data harmonization workshops.
Uneven staff capacity across regions	Create regional capacity-building programs and mentorship networks. Offer certification-based data management training.
Non-reporting & biased reporting	Incentivize timely reporting with performance dashboards and feedback loops. Establish independent verification teams.
Poor interoperability with partners	Develop APIs and middleware for data exchange between systems. Use open standards like FHIR or OpenHIE architecture.
Language / translation / context loss	Standardize multilingual data entry forms and metadata. Involve local translators in data validation and verification stages.

References

- Bakken, S., Dykes, P. C., Collins Rosetti, S., & Ozbolt, J. G. (2021). Patient-Centered Care Systems. In Shortliffe, E. H. & Cimino, J. J. (Eds.) & Chiang, M. F. (Co-Ed.). (2021). *Biomedical informatics: Computer applications in healthcare and biomedicine*. (5th ed.). New York: Springer.
- Cahill, M., Cleary, B. J., & Cullinan, S. (2025). The influence of electronic health record design on usability and medication safety: Systematic review. *BMC Health Services Research*, 25(1), 31. <https://doi.org/10.1186/s12913-024-12060-2>
- CCD Health. (2025, February 2). *How healthcare data quality impacts patient outcomes*. <https://ccdcare.com/resource-center/healthcare-data-quality/>
- Cruz-Correia, R., Rodrigues, P., Freitas, A., Almeida, F., Chen, R., & Costa-Pereira, A. (2009). Data quality and integration issues in electronic health records. In V. Hristidis (Ed.), *Information discovery on electronic health records* (pp. 55–95). CRC Press. <https://doi.org/10.1201/9781420090413-c4>
- Felt-Lisk, S., Johnson, L., Fleming, C., Shapiro, R., & Natzke, B. (2010). Toward understanding EHR use in small physician practices. *Health Care Financing Review*, 31(1), 11–22.
- Gazi, T. (2020). Data to the rescue: How humanitarian aid NGOs should collect information based on the GDPR. *Journal of International Humanitarian Action* 5(9). <https://doi.org/10.1186/s41018-020-00078-0>
- Gharawi, M.A., Pardo, T.A., & Guerrero, S. (2009). Issues and strategies for conducting cross-national e-government comparative research. *International Conference on Theory and Practice of Electronic Governance*, 163–170. <https://doi.org/10.1145/1693042.1693076>
- Madandola, O. O., Bjarnadottir, R. I., Yao, Y., Ansell, M., Dos Santos, F., Cho, H., Dunn Lopez, K., Macieira, T. G. R., & Keenan, G. M. (2023). The relationship between electronic health records user interface features and data quality of patient clinical information:

Group 3

B585 – Biomedical Analytics - Assignment 3

An integrative review. *Journal of the American Medical Informatics Association*, 31(1),

240–255. <https://doi.org/10.1093/jamia/ocad188>

Masefield, S.C., Megaw, A., Barlow, M., White, P., C., L., Altink, H., & Grugel, J. (2020).

Repurposing NGO data for better research outcomes: A scoping review of the use and secondary analysis of NGO data in health policy and systems research. *Health Research Policy and Systems*, 18(63). <https://doi.org/10.1186/s12961-020-00577-x>

National Center for Health Statistics. (2024, December 13). *Electronic medical records/electronic health records (EMRs/EHRs)*. National Center for Health Statistics,

Centers for Disease Control and Prevention.

<https://www.cdc.gov/nchs/fastats/electronic-medical-records.htm>

Patel, V. L., Kaufman, D. R., & Kannampallil, T. (2021). Human-computer interaction, usability,

and workflow. In Shortliffe, E. H. & Cimino, J. J. (Eds.) & Chiang, M. F. (Co-Ed.). (2021).

Biomedical informatics: Computer applications in healthcare and biomedicine. (5th ed.,

pp. 153-176). New York: Springer.

van Zyl, C., Badenhorst, M., Hanekom, S., & Heine, M. (2021). Unravelling 'low-resource

settings': A systematic scoping review with qualitative content analysis. *BMJ Global Health*, 6(6), e005190. <https://doi.org/10.1136/bmjgh-2021-005190>

Whalen, C. J., Donnell, D., & Tartakovsky, M. (2014). Supporting research sites in resource-

limited settings: Challenges in implementing information technology infrastructure.

Journal of Acquired Immune Deficiency Syndrome, 65, S44–S49.

<https://doi.org/10.1097/QAI.0000000000000039>