

Data Ordinality: An Important Concept in Reducing Medical and Research Errors

This manuscript ([permalink](#)) was automatically generated from [fmauire/data_ordinality_paper@20551e0](#) on October 7, 2019.

Authors

- **Finlay Maguire**

 [0000-0002-1203-9514](#) ·  [fmauire](#) ·  [fmauire](#)

Faculty of Computer Science, Dalhousie University · Funded by Donald Hill Family Fellowship in Computer Science

- **Michael A. Graven**

Department of Paediatrics, Faculty of Medicine, Dalhousie University

- **Noni E. MacDonald**

Department of Paediatrics, Faculty of Medicine, Dalhousie University

Data Ordinality

//% of medical errors that are related to transcription //The majority of medical research is retrospective //The more times data is copied the more errors //Different types of data have different error rates //Easier to avoid errors than fix them //Humans are always going to make errors

Adverse effects of medical treatment are responsible for 2.8% of all deaths in the US [1]. ICU patients experience an 1.7 medical errors a day of which 29% were errors classed as having the potential to cause severe harm or death [2] A major source of these adverse outcomes is errors in the transmission and communication of information between healthcare workers (HCW). For example, transcription errors (misinterpretation of written/spoken orders and mistakes copying prescriptions) are responsible for 2-14% of medication errors across a range of contexts [3]. Humans, even very well trained ones, are fallible. This means human error is inevitable.

While errors in patient care can have disastrous consequences for that patient, these errors can and do propagate into clinical research. In some cases these errors distort the results of clinical trials leading harming potentially huge numbers of patients by leading to incorrect changes in the standards of care. Unfortunately, these errors aren't evenly and randomly distributed across the entire dataset with some bits of data having discrepancy rates as high as 26.9% [4]. With some analyses showing that an error on a page greatly increased the risk of another error appearing on the same page [4].

The more times a piece of data is copied the more likely it is to have errors introduced. An intuitive example of this effect is that of the childhood "whisper game". In this, a message is whispered from person to person down a line of people. Upon the last person in the line speaking the message aloud the message is typically drastically different from the original message. Each person listening and repeating this message allows the original information to become increasingly distorted and warped. Medical research is particularly affected by this especially as a significant quantity of all clinical research involves retrospective review of patient charts (1/4 of emergency medicine research studies [5]).

To help researchers and clinicians understand and conceptualise this we present the principle of "data ordinality".

Ordinality is a term used in maths to denote the order of objects within a sequence, e.g., 1st, 2nd, 3rd. We can apply this idea to data, by using it describe how many times a piece of data has been manually entered or written down. For example, 1st order data would be data that has been directly extracted or printed out from a device such as a digital thermometer.

When this temperature is written down by a HCW into a patient's chart it would now be 2nd order data. Finally, when a researcher extracts this temperature from the chart it is now 3rd order data. In our intuitive example, data ordinality would be the number of whisper events in the game of telephone/whisper. Errors increase as data ordinality increases, much as the distortion of the message becomes greater the more people are playing the whisper game.

This is an analogous generalisation of the concept of historiographical primary, secondary, and tertiary sources.

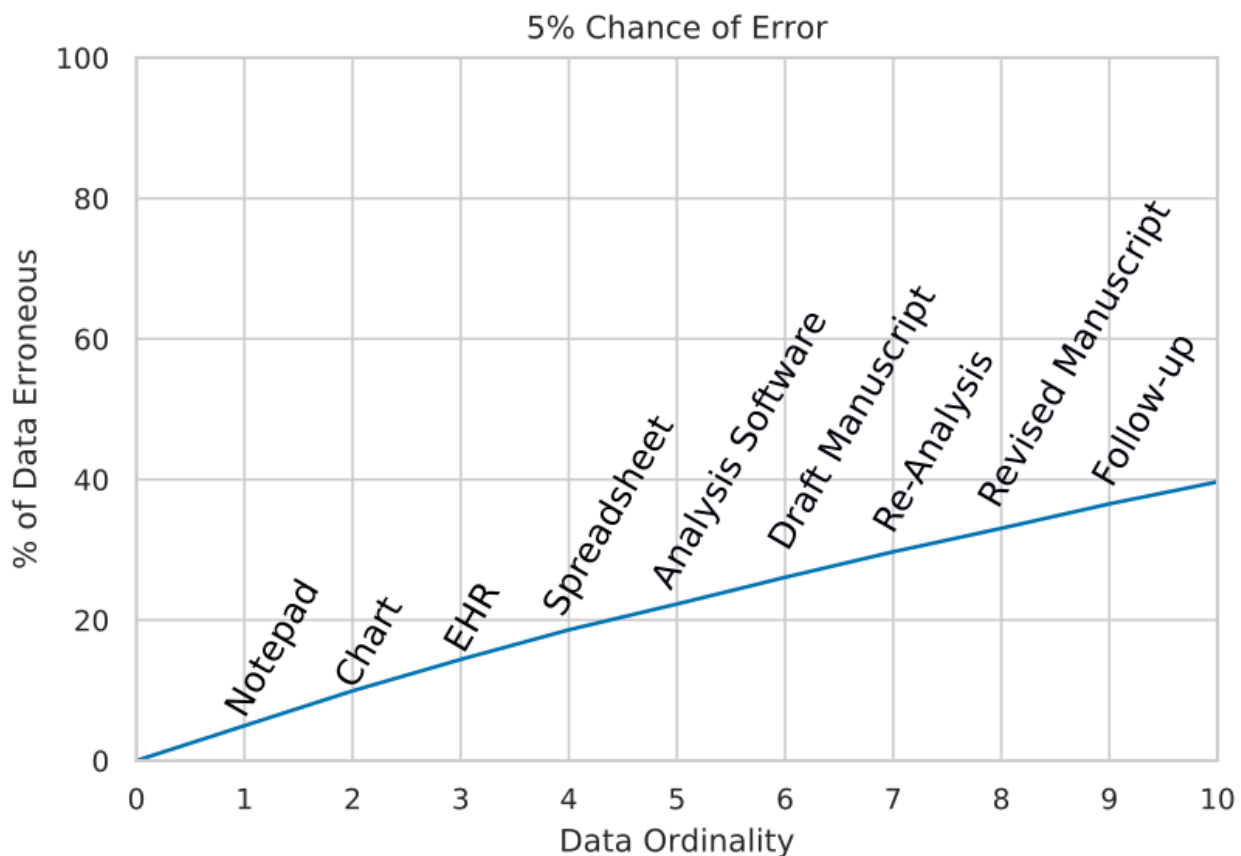


Figure 1: Propagation of Error with Data Ordinality. Assuming an arbitrary uniform 5% chance of error in the copying of each data-point at each stage of copying. In reality different pieces of data and different stages will have drastically different error rates.

One way to reduce data ordinality is the adoption of clinical information technology (IT) such as electronic health records, electronic prescription systems etc. The adoption of these systems have been shown to reduce hospital mortality, complications, and costs [6]. However, just because data has been entered into a computer does not mean its ordinality cannot still increase. A dataset can be sent to one researcher, who then makes some accidental changes due to either user error or computer glitches (partial data transfer, bugs, automated reformatting).

Typically, it is always easier and cheaper to prevent errors and adverse situations than to retroactively attempt to fix them.

A pervasive example of this in the life sciences is the mangling of gene names through automated formatting of them as dates by Excel [8]. These mangled names are then uploaded to central databases where they continue to proliferate as other researchers make use of them.

Data ordinality can be controlled by having a single authoritative source copy of any dataset with restricted audited ability to change it. This dataset and any updates to it should be tracked using a version control system and kept thoroughly backed-up in order to maintain data integrity. Indeed, these are a requirements enumerated in many national laws and regulations governing clinical trials and pharmaceutical manufacture (e.g. US Federal Drug Association's 21CFR11).

Ultimately, we believe that the concept of data ordinality is a useful mental framework that clinicians should be mindful of in both practice and research.

References

1. Association of Adverse Effects of Medical Treatment With Mortality in the United States

Jacob E. Sunshine, Nicholas Meo, Nicholas J. Kassebaum, Michael L. Collison, Ali H. Mokdad, Mohsen Naghavi

JAMA Network Open (2019-01-18) <https://doi.org/gf9gds>

DOI: [10.1001/jamanetworkopen.2018.7041](https://doi.org/10.1001/jamanetworkopen.2018.7041) · PMID: [30657530](https://pubmed.ncbi.nlm.nih.gov/30657530/) · PMCID: [PMC6484545](https://pubmed.ncbi.nlm.nih.gov/PMC6484545/)

2. A look into the nature and causes of human errors in the intensive care unit

Yoel Donchin, Daniel Gopher, Miriam Olin, Yehuda Badihi, Michal RNB Biesky, Charles L. Sprung, Ruven Pizov, Shamay Cotev

Critical Care Medicine (1995-02) <https://doi.org/fmndxj>

DOI: [10.1097/00003246-199502000-00015](https://doi.org/10.1097/00003246-199502000-00015) · PMID: [7867355](https://pubmed.ncbi.nlm.nih.gov/7867355/)

3. Drug-Related Problems in Hospitals

Anita Kr??henb??hl-Melcher, Raymond Schlienger, Markus Lampert, Manuel Haschke, J??rgen Drewe, Stephan Kr??henb??hl

Drug Safety (2007) <https://doi.org/bg3z3z>

DOI: [10.2165/00002018-200730050-00003](https://doi.org/10.2165/00002018-200730050-00003) · PMID: [17472418](https://pubmed.ncbi.nlm.nih.gov/17472418/)

4. Analysis of data errors in clinical research databases.

Saveli I Goldberg, Andrzej Niemierko, Alexander Turchin

AMIA ... Annual Symposium proceedings. AMIA Symposium (2008-11-06)

<https://www.ncbi.nlm.nih.gov/pubmed/18998889>

PMID: [18998889](https://pubmed.ncbi.nlm.nih.gov/18998889/) · PMCID: [PMC2656002](https://pubmed.ncbi.nlm.nih.gov/PMC2656002/)

5. Chart reviews in emergency medicine research: Where are the methods?

EH Gilbert, SR Lowenstein, J Koziol-McLain, DC Barta, J Steiner

Annals of emergency medicine (1996-03) <https://www.ncbi.nlm.nih.gov/pubmed/8599488>

DOI: [10.1016/s0196-0644\(96\)70264-0](https://doi.org/10.1016/s0196-0644(96)70264-0) · PMID: [8599488](https://pubmed.ncbi.nlm.nih.gov/8599488/)

6. Clinical Information Technologies and Inpatient Outcomes

Ruben Amarasingham, Laura Plantinga, Marie Diener-West, Darrell J. Gaskin, Neil R. Powe

Archives of Internal Medicine (2009-01-26) <https://doi.org/bzmhbp>

DOI: [10.1001/archinternmed.2008.520](https://doi.org/10.1001/archinternmed.2008.520) · PMID: [19171805](https://pubmed.ncbi.nlm.nih.gov/19171805/)

7. Does the Leapfrog program help identify high-quality hospitals?

Ashish K Jha, E John Orav, Abigail B Ridgway, Jie Zheng, Arnold M Epstein

Joint Commission journal on quality and patient safety (2008-06)

<https://www.ncbi.nlm.nih.gov/pubmed/18595377>

PMID: [18595377](https://pubmed.ncbi.nlm.nih.gov/18595377/)

8. Gene name errors are widespread in the scientific literature

Mark Ziemann, Yotam Eren, Assam El-Osta

Genome Biology (2016-08-23) <https://doi.org/bqt3>

DOI: [10.1186/s13059-016-1044-7](https://doi.org/10.1186/s13059-016-1044-7) · PMID: [27552985](https://pubmed.ncbi.nlm.nih.gov/27552985/) · PMCID: [PMC4994289](https://pubmed.ncbi.nlm.nih.gov/PMC4994289/)

9.

Barry R Zeeberg, Joseph Riss, David W Kane, Kimberly J Bussey, Edward Uchio, W Marston Linehan, J Carl Barrett, John N Weinstein

BMC Bioinformatics (2004) <https://doi.org/fcvtjq>

DOI: [10.1186/1471-2105-5-80](https://doi.org/10.1186/1471-2105-5-80) · PMID: [15214961](https://pubmed.ncbi.nlm.nih.gov/15214961/) · PMCID: [PMC459209](https://pubmed.ncbi.nlm.nih.gov/PMC459209/)

