



Opinion

The power of small data for personalized cancer care

Junqiang Zhao

School of Nursing, Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada



Big data has been flourishing in the healthcare field in those years and has exerted a profound impact on the way how nursing practice and research are conducted. While it has numerous and to some extent, irreplaceable advantages to transform healthcare, there are critiques that pure big data approach is insufficient to achieve personalized care promise due to the complex, multi-causal, and idiosyncratic nature of human health conditions.^{1,2} In this article, I advocate for a small data paradigm for personalized cancer care, and more broadly, nursing research and practice. I argue that small data has large potential in building causal explanations and facilitating rapid learning. Small and big data should be used in an integrated and complementary fashion to promote the personalized practice.

Small data has been defined in various ways in different disciplines. Researchers in the informatics field compared it with big data from the features of volume, variety, and velocity and considered small data as small in volume, narrow in variety, and slow in velocity.³ In the business field, the author of the book *Small Data - The tiny clues that uncover huge trends* Martin Lindstrom defined small data as those seemingly insignificant observations we make in people's lives that lead to an understanding of causations.⁴ In healthcare, Estrin viewed small data as the $n = me$ data that we generate every day for ourselves.⁵ In this article, I turn to the definition by Hekler et al and consider small data as "the rigorous use of data by and for a specific N-of-1 unit (ie., a single person, clinic, hospital, healthcare system, community, city, etc.) to facilitate improved individual-level description, prediction and, ultimately, control for that specific unit".² Here the word "small" does not necessarily mean that the data volume is small but that the data is collected from and being used for a single unit. While big data aims to collect information from a set of individuals to improve the description and prediction for other individuals, small data collects individual data for that individual.^{2,5} It should be noted that while small data has been dominantly discussed in artificial intelligence, machine learning, and informatics fields, it is not the focus of this article to delve into these aspects. I consider small data as a research paradigm or orientation that guides our deep understanding of complex nursing, healthcare, and social phenomena rather than a type of analytical method.

In this article, I start with an ontological argument for the small data approach and suggest that small data, consistent with an ontological

position of causal dispositionalism, has a larger potential to promote personalized care than the big data approach underpinned by empiricism and Humeanism. Next, I raise two advantages of small data in healthcare research and practice: building causal explanations and facilitating rapid learning. Then, I illustrate two promising research methods that align well with the small data paradigm: the case study method in social science and the N-of-1 trial in natural science. After that, I discuss the possible ways to integrate big and small data to promote personalized care. Lastly, I propose several implications of small data for personalized cancer care.

Small data, causal dispositionalism, and personalized care

Causation is one of the basic features of reality and the main foundation upon which science is based.⁶ Big data and small data can be viewed as being underpinned by different ontological positions of causation. Big data, akin to evidence-based medicine, has a solid foundation in empiricism. Empiricism reduces complex human conditions into its compositions and builds causal relations and universal laws through correlations between variables. It is premised on the idea that what works at the population level should guide the decision for individuals.⁷ The position of causation under this paradigm is called Humeanism. Humeanism holds four main propositions on causation, namely (1) constant conjunction: causality is the perfect correlation between a cause and its effect; every time the cause happens, the effect follows; (2) temporal priority: the cause always happens before the effect; (3) contiguity: the cause and effect must be contiguous in space and time; and (4) the same cause always produces the same effect, and the same effect never arises but from the same cause.^{6,8,9} For example, it is now well-accepted that smoking is the leading cause of lung cancer. From the Humeanism perspective, the reason why smoking causes lung cancer is that the epidemiological data show that heavy smokers often (constant conjunction) suffer from lung cancer after a period of time (temporal priority and contiguity), and when we trace back the history of patients with lung cancer, many of them smoke. However, Humeanism faces serious critiques in building real causation due to its failure to distinguish correlation from causation. The same cause does not necessarily lead to

E-mail address: jzhao126@uottawa.ca.<https://doi.org/10.1016/j.apjon.2022.100084>

Received 25 April 2022; Accepted 10 May 2022

2347-5625/© 2022 The Author(s). Published by Elsevier Inc. on behalf of Asian Oncology Nursing Society. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

the same effect across different contexts—context matters in the causal process. Causes and effects are not linear relations. One cause might have multiple intended and unintended effects; one effect might have multiple interacting causes.^{6,8,9} Therefore, the pure big data approach underpinned by empiricism and Humeanism can hardly achieve personalized care promise because of its limitations on external validity, the ecological fallacy,¹⁰ and the non-linearity, complexity, and multi-causality of human health conditions.^{11,12}

Small data, on the other hand, can be viewed as consistent with an ontological position of causal dispositionalism. Different from Humeanism, dispositionalism holds that causation happens in a particular case and does not require repetition. Disposition is the causal power, abilities, capacities, or the intrinsic properties of things that can exist unmanifested. It is a tendency toward effects—causality is the result of complex interaction among multiple dispositions.^{6,8,10} The reason why smoking leads to lung cancer is not the correlation between these two phenomena, although such correlation might lead to our understanding of causation. The real reason is that the cigarette smoke we breathe in contains chemicals that have the disposition to cause genetic mutation in the lung cells. From a dispositionalist view, to establish a causal relationship between an intervention and an effect is to understand whether the intervention has dispositions that interact with other dispositions to produce the effects.¹³ The same disposition tends to produce different effects depending on which dispositions it interacts with. For example, Ibuprofen, as one type of nonsteroidal anti-inflammatory drugs (NSAIDs), has the disposition to relieve cancer pain. However, it only works with patients who suffer from mild pain and have the disposition to react to NSAIDs. When patients suffer from severe pain, neuropathic pain, or bone metastases pain, they no longer have the disposition to react to NSAIDs, thus ibuprofen seldom works in these circumstances. Dispositionalism embraces the complex and multifactorial nature of causation, individual variations, and context sensitivities. It argues that causation is singular and each patient is unique and thus one size does not fit all.¹⁰ The small data approach, collecting rich data about the specific N-of-1 unit and appreciating the heterogeneity of individuals, has large potential in building causations for individuals (described below) and thus promoting personalized care.

Advantages of small data paradigm

Small data for building causal explanations

Small data has large potential in building causal explanations. Under dispositionalism, causation is understood as the intrinsic dispositions of things and thus can not be captured from the pure empirical facts. Researchers need to go beyond the investigation of whether and how often things happen through statistical analysis (ie. empirical level) and delve into the theoretical knowledge on why and how things happen or do not happen (ie. causal explanations). The small data approach, which can also be considered an “intensive research” approach,¹⁴ collects different types of qualitative, quantitative, and longitudinal data, such as the medical examinations, lab analyses, medical history, and patient narratives on lived experiences, around a particular case or a few cases. It enables a deep understanding of the surrounding contexts and their interactions with an individual to build context-sensitive explanations. To put it in another way, it helps to gain insight into how multiple dispositions surrounding the individual interact to cause effects.

Social scientists from the critical realism tradition argue that researchers can use retroductive reasoning to build causal explanations qualitatively by conducting intensive research (ie. small data research). Retroduction is to advance from empirical observations of events and identify the causal power of structures by using a combination of induction, deduction, and abduction logics.¹⁴ They suggest that the deep interrogation of pathological cases, extreme cases, and comparative cases are helpful to identify causal explanations.^{14,15} Researchers can use a five-stage framework in an intensive study to build causal explanation: (1) descriptions—describe the complex phenomena that we intend to

explain; (2) analytical resolution—distinguish the various components, aspects, dimensions, and levels of analysis and clarify what we aim to explain; (3) abduction—interpret and redescribe the components of the phenomena using theoretical frameworks for possible explanations; (4) retroduction—find answers to the question: “*What is fundamentally constitutive for the structures and what mechanisms are related to these structures?*” by using the aforementioned case study strategies; and (5) comparisons and contextualization—evaluate the explanatory power of different mechanisms using theoretical and empirical approaches and understand the context-sensitivities.^{14–17} These five stages are often intertwined, rather than in strictly chronological order, to build causal explanations from the concrete (stage 1) to abstract (stage 2–4) and then back to the concrete (stage 5).¹⁴ Currently, very limited small data studies have been done that draw on critical realism to build causal understandings of nursing phenomena, whereas it has become increasingly used in education,¹⁸ information science,¹⁹ and management fields.²⁰

Causal dispositionalism philosophers suggested a three-stage iterative causal discovery process, including (1) observing phenomenon; (2) hypothesizing causation; and (3) establishing causation.¹³ The small data approach runs through all these stages. Based on the proposed process, we can start from an observed phenomenon and gain a deep understanding of this phenomenon in its context in sufficient details through case studies. We then enter stage 2 to hypothesize causation. Building on existing knowledge, one can test the hypothesis through experiments, such as the N-of-1 trials, and further observations, such as patient narratives. After that, we can try to establish causation by observing cases of causal failure, in which dispositions do not manifest in an expected way. We may have a clear understanding of causation at this stage, or we need to go back to stage 1 to continue the process.¹³ Anjum et al, in a recent publication, proposed a *Dx3* approach for assessing whether a particular medicine has or could have caused a certain adverse event by using small data (ie. individual case safety reports).²¹ This approach argues that three types of dispositions—the drug disposition, the predisposition of the patient taking the drug (vulnerability), and the disposition of the patient–drug interaction (mutuality)—can be qualitatively evaluated to understand causality in pharmacovigilance. They also provide guidance, checklists, and examples on how to conduct such causal assessments.²¹ It should be noted that here I aim to emphasize the significant, and at some points, irreplaceable role of small data throughout the causation building process. That does not mean other research methods, such as randomized controlled trials, cohort studies, case-control studies, and lab models, contribute little to the process. These methods all contribute to the causal process in one way or another, especially in the testing of causal hypotheses.¹³

Small data for rapid learning

Small data can also promote rapid learning. While the big data approach uses data from a group of individuals to produce transportable knowledge primarily for other individuals, the defining feature of small data is that the data are collected from and primarily used by individuals for their own purposes. We become the consumers of our own data in the small data paradigm.^{2,5} Similar to local quality improvement projects, small data becomes advantageous for rapid learning at the N-of-1 unit. This unit can learn from and respond to the data findings rapidly. When the N-of-1 unit is a single patient, we can adjust treatment and care plans for this individual agilely based on his/her own process and outcome data; when the N-of-1 unit is one healthcare organization, the small data becomes valuable for advancing a learning healthcare organization where “knowledge generation processes are embedded in daily practice to produce continual improvement in care”.²² Such an advantage of small data is named *reinforcement learning* in artificial intelligence—“an agent learns how to interact with its environment via trial and error” (p. 8).²³

The big data approach uses population-level data to inform individual-level practice through clustering people based on variables (ie. statistical generalization), it is challenging for rapid learning due to

the impact of local complex and dynamic contexts. The small data approach, on the other hand, produces theoretical knowledge on causal mechanisms and informs population-level practice by clustering people with similar mechanisms (ie. theoretical generalization).^{1,2} When the causal mechanisms identified from small data fail to be applied to another individual, we can instantly cease its generalization and reflect on the explanatory power of the mechanisms and contextual differences. The N-of-1 unit is sufficient to trigger reflection and learning.² Such an advantage of small data is named *transfer learning* in artificial intelligence—it “works by first learning how to perform a task in a setting where data is abundant, then transferring what it has learned there to a task where much less data is available” (p.6).²³

Small data-oriented research methods

In this section, I describe two promising research methods that are small data-oriented, namely the case study method in social science and the N-of-1 trial in natural science.

Case study

A case study is a commonly used research method in social science. It is the study of a real-life, contemporary case or cases over time through detailed and in-depth data collection involving multiple sources of information, such as observations, interviews, and documents.²⁴ A case is a bounded system and can be defined or described within certain parameters. While the case study approach can be used for exploratory, descriptive, and explanatory purposes, its potential to build causal explanations is largely untapped. Case study has been advised as a promising method to explain complex systems as it can capture the complex and dynamic interaction among system agents through longitudinal and cross-case analysis.^{25,26} A well-conducted empirical case study allows for an understanding of the dynamic and evolving influence of context on complex system-level interventions, thus generating theoretical knowledge on what interventions work in what context to achieve desired effects.²⁶ The analytical method of qualitative comparative analysis has become popular in case studies to identify the configurations of causal conditions for an intended outcome.²⁷ A recent meta-narrative review of case study approaches identified four broad research traditions in evaluating complex interventions: (1) developing and testing complex interventions in healthcare; (2) analyzing changes in organizations; (3) undertaking realist evaluations; and (4) studying complex change naturalistically.²⁸

N-of-1 trial

N-of-1 trials are prospective single-patient trials with multiple cross-over periods between interventions and comparators.²⁹ This type of trial uses key methodological components from traditional group-based clinical trials to evaluate the effectiveness of treatment in a single patient. This design is particularly useful for situations that can not accommodate large-scale trials, such as patients with rare diseases, comorbid conditions, or using concurrent therapies, and is more appropriate for patients with stable chronic conditions.³⁰ The ultimate goal of an N-of-1 trial is to determine the optimal evidence-based treatment for an individual patient. So, a major advantage of the one-person trial is that the effectiveness of treatments is vetted for the participants themselves rather than benefiting other people which is the goal of traditional trials.³¹ N-of-1 trials are also good tools for establishing difference-making (ie. the philosophical idea that causes tend to make a difference; hence, if one can discover the difference-makers, then it can be a good indicator of causes⁶) in a unique context, and therefore, have the epistemic advantages for establishing causation.¹³ The aggregated results of multiple N-of-1 trials have the potential to inform population-level treatment.³¹ In 2015, Nature published a commentary to call for the scaling up of N-of-1 trials in everyday clinical practice to promote precision medicine and personalized care.³¹

Complementary use of big and small data

Big data has become an irreversible trend in current healthcare. Rather than devaluing big data contributions and prioritizing small data, it is more meaningful to integrate them at the doctor–patient encounter to achieve the personalized care promise.^{1,2,32} While big data generates scientific knowledge from statistical analysis of variables, small data produces tacit knowledge from patient narratives. Big data provides an understanding of general patterns on correlations, whereas small data gains insights into causations. Small data can build on big data findings and refine them in an individual patient to understand the context sensitivities. Big data become established by collecting high-quality small data. Sacristan and Dilla suggested that electronic medical records were places where small and big data could perfectly intersect. Small data is collected at each doctor-patient encounter and uploaded to the electronic medical records. Big data becomes established when those complete and standardized small data are summed up.³²

The complementary use of the big and small data approach, as I propose, can have two common forms: the bottom-up approach and the top-down approach. The bottom-up approach collects high-quality small data to build big data and informs not only individual care but also policy-making. This approach has been vividly described in a case study of fundamental nursing care.³³ Drawing on the complexity science, Conroy et al described in detail the development of data matrices through analyzing patient narratives to quantify and evaluate fundamental care in a single patient. Building on the small data, they are planning to standardize the procedures and use computer algorithms and digital technology to establish big data. The complementary use of the small and big data, in this case, can not only inform the fundamental care at the individual level but also help reveal predictive patterns of care and build an early warning system to prevent the deterioration of care standards.³³ The top-down approach draws on findings from existing big data and fine-tunes them in individual cases. For example, Zhu et al described the use of big data-based network analysis to identify the symptom patterns in a population to inform symptom care³⁴; Zhao argued that to achieve personalized symptom care, we should turn to narratives and medical records to refine and consolidate the symptom patterns in an individual patient.¹

Implications of small data for personalized cancer care

While small data is a relatively new concept, the use of small data-oriented methods has a long history already. Nevertheless, it is still underused in the nursing field and its power in promoting personalized nursing care is underestimated. The small data approach has several implications for personalized cancer care, and more broadly, nursing practice: (1) Integrate clinical practice with nursing research—transforming everyday cancer care practice into solid individual-level data collection³¹ and improving the quality of data entered in the electronic medical records.³² (2) Embrace patient narratives—listen to stories of patients with cancer to understand their bio-physio-social contexts, which will facilitate our understandings of the underlying causes of patient symptoms.¹ (3) Employ a complex science lens to uncover the interrelationships among patient's cancer symptoms and, more broadly, health conditions.^{1,33} Human body is a complex adaptive system which requires us to take a holistic rather than a reductionist view on patients' conditions. (4) Interdisciplinary efforts to understand the causation behind cancer symptoms to inform nursing practice. (5) Perform shared decision making with patients on care plans and be ready to adjust them agilely based on rapid learning.

Conclusions

In this article, I distinguished the ontological positions of small and big data approaches and illustrated two unique advantages of small data: building causation and facilitating rapid learning. After that, I described two commonly used methods that are small data-oriented: case study and

N-of-1 trial. Small data and big data should be used in a complementary fashion to promote personalized care. Such integration is promising to facilitate the close collaboration between researchers (the big data approach) and healthcare professionals (the small data approach) and is likely to translate research evidence into practice timely.¹ Lastly, I proposed several implications of small data for personalized cancer care practice.

Declaration of competing interest

None declared.

References

- Zhao J. Is there such a thing as a symptom cluster: the paradigm shift in symptom science requires a philosophical reflection. *Asia Pac J Oncol Nurs*. 2022;9(5):100056. <https://doi.org/10.1016/j.apjon.2022.03.008>.
- Hekler EB, Klasnja P, Chevance G, Golaszewski NM, Lewis D, Sim I. Why we need a small data paradigm. *BMC Med*. 2019;17(1):133. <https://doi.org/10.1186/s12916-019-1366-x>.
- Kitchin R, Lauriault TP. Small data in the era of big data. *Geojournal*. 2015;80(4):463–475. <https://doi.org/10.1007/s10708-014-9601-7>.
- Lindstrom M. *Small Data: The Tiny Clues that Uncover Huge Trends*. St. Martin: s Press; 2016.
- Estrin D. Small data, where n = me. *Commun ACM*. 2014;57(4):32–34. <https://doi.org/10.1145/2580944>.
- Anjum RL, Mumford S. *Causation in Science and the Methods of Scientific Discovery*. USA: Oxford University Press; 2018.
- Hickey S, Hickey A, Noriega L. The failure of evidence-based medicine? *Eur J Pers Cent Healthc*. 2013;1(1):69–79.
- Anjum RL, Copeland S, Rocca E. *Rethinking Causality, Complexity and Evidence for the Unique Patient: A Causehealth Resource for Healthcare Professionals and the Clinical Encounter*. Springer Nature; 2020.
- Kerry R, Eriksen TE, Lie SA, Mumford SD, Anjum RL. Causation and evidence-based practice: an ontological review. *J Eval Clin Pract*. 2012;18(5):1006–1012. <https://doi.org/10.1111/j.1365-2753.2012.01908.x>.
- Anjum RL. Evidence based or person centered? An ontological debate. *Eur J Pers Cent Healthc*. 2016;4(2).
- Sturmberg JP. Evidence-based medicine-Not a panacea for the problems of a complex adaptive world. *J Eval Clin Pract*. 2019;25(5):706–716. <https://doi.org/10.1111/jep.13122>.
- Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g3725. <https://doi.org/10.1136/bmj.g3725>.
- Rocca E, Anjum RL. Causal evidence and dispositions in medicine and public health. *Int J Environ Res Publ Health*. 2020;17(6):1813. <https://doi.org/10.3390/ijerph17061813>.
- Danermark B, Ekstrom M, Jakobsen L. *Explaining Society: An Introduction to Critical Realism in the Social Sciences*. 2nd ed. Routledge; 2019.
- Meyer SB, Lunnay B. The application of abductive and retroductive inference for the design and analysis of theory-driven sociological research. *Socio Res Online*. 2013;18(1):86–96.
- Thapa D, Omland HO. Four steps to identify mechanisms of ICT4D: a critical realism-based methodology. *Electron J Inf Syst Dev Ctries*. 2018;84(6), e12054. <https://doi.org/10.1002/isd2.12054>.
- Bygstad B, Munkvold BE, Volkoff O. Identifying generative mechanisms through affordances: a framework for critical realist data analysis. *J Inf Technol*. 2016;31(1):83–96.
- Stutchbury K. Critical realism: an explanatory framework for small-scale qualitative studies or an 'unhelpful edifice'. *Int J Res Method Educ*. 2021:1–16.
- Mingers J, Standing C. Why things happen—Developing the critical realist view of causal mechanisms. *Inf Organ*. 2017;27(3):171–189.
- Williams CK, Karahanna E. Causal explanation in the coordinating process: a critical realist case study of federated IT governance structures. *MIS Q*. 2013:933–964.
- Anjum RL, Chandler RE, Rocca E. Dispositions and causality assessment in pharmacovigilance: proposing the Dx3 approach for assessing causality with small data sets. *Pharmaceut Med*. 2022:1–9.
- Olsen L, Aisner D, McGinnis JM. *The Learning Healthcare System: Workshop Summary*. 2007.
- Chahal H, Toner H, Rahkovsky I. *Small Data's Big AI Potential*. 2021.
- Creswell JW, Poth CN. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Sage publications; 2016.
- Anderson RA, Crabtree BF, Steele DJ, McDaniel Jr RR. Case study research: the view from complexity science. *Qual Health Res*. 2005;15(5):669–685. <https://doi.org/10.1177/1049732305275208>.
- Paparini S, Green J, Papoutsis C, et al. Case study research for better evaluations of complex interventions: rationale and challenges. *BMC Med*. 2020;18(1):301. <https://doi.org/10.1186/s12916-020-01777-6>.
- Rihoux B, Ragin CC. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Sage Publications; 2008.
- Paparini S, Papoutsis C, Murdoch J, et al. Evaluating complex interventions in context: systematic, meta-narrative review of case study approaches. *BMC Med Res Methodol*. 2021;21(1):225. <https://doi.org/10.1186/s12874-021-01418-3>.
- Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. *Ann Intern Med*. Feb 15 1990;112(4):293–299. <https://doi.org/10.7326/0003-4819-112-4-293>.
- Vohra S, Shamseer L, Sampson M, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ*. May 14 2015;350:h1738. <https://doi.org/10.1136/bmj.h1738>.
- Schork NJ. Personalized medicine: time for one-person trials. *Nature*. 2015;520(7549):609–611.
- Sacristán JA, Dilla T. No big data without small data: learning health care systems begin and end with the individual patient. *J Eval Clin Pract*. Dec 2015;21(6):1014–1017. <https://doi.org/10.1111/jep.12350>.
- Conroy T, Pinero de Plaza MA, Mudd A, Mitchell M, Kitson A. Measuring fundamental care using complexity science: a descriptive case study of a methodological innovation. *J Clin Nurs*. 2021. <https://doi.org/10.1111/jocn.15905>.
- Zhu Z, Xing W, Hu Y, Wu B, So WK. Paradigm shift: moving from symptom clusters to symptom networks. *Asia Pac J Oncol Nurs*. 2022;9(1):5–6.