



Commentary

Reproducibility and Scientific Integrity of Big Data Research in Urban Public Health and Digital Epidemiology: A Call to Action

Ana Cecilia Quiroga Gutierrez ^{1,*}, Daniel J. Lindegger ^{2,†}, Ala Taji Heravi ³, Thomas Stojanov ⁴,
Martin Sykora ⁵, Suzanne Elayan ⁵, Stephen J. Mooney ⁶, John A. Naslund ⁷, Marta Fadda ⁸
and Oliver Gruebner ^{9,10}

¹ Department of Health Sciences and Medicine, University of Lucerne, 6002 Luzern, Switzerland

² Institute of Global Health, University of Geneva, 1211 Geneva, Switzerland

³ CLEAR Methods Center, Department of Clinical Research, Division of Clinical Epidemiology,
University Hospital Basel and University of Basel, 4031 Basel, Switzerland

⁴ Department of Orthopaedic Surgery and Traumatology, University Hospital of Basel, 4031 Basel, Switzerland

⁵ School of Business and Economics, Centre for Information Management, Loughborough University,
Loughborough LE11 3TU, UK

⁶ Department of Epidemiology, University of Washington, Seattle, WA 98195, USA

⁷ Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA 02115, USA

⁸ Institute of Public Health, Università Della Svizzera Italiana, 6900 Lugano, Switzerland

⁹ Epidemiology, Biostatistics and Prevention Institute, University of Zurich, 8001 Zurich, Switzerland

¹⁰ Department of Geography, University of Zurich, 8057 Zurich, Switzerland

* Correspondence: ana.quiroga@unilu.ch

† These authors contributed equally to this work.



Citation: Quiroga Gutierrez, A.C.; Lindegger, D.J.; Taji Heravi, A.; Stojanov, T.; Sykora, M.; Elayan, S.; Mooney, S.J.; Naslund, J.A.; Fadda, M.; Gruebner, O. Reproducibility and Scientific Integrity of Big Data Research in Urban Public Health and Digital Epidemiology: A Call to Action. *Int. J. Environ. Res. Public Health* **2023**, *20*, 1473. <https://doi.org/10.3390/ijerph20021473>

Academic Editor: William

Douglas Evans

Received: 18 November 2022

Revised: 31 December 2022

Accepted: 11 January 2023

Published: 13 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The emergence of big data science presents a unique opportunity to improve public-health research practices. Because working with big data is inherently complex, big data research must be clear and transparent to avoid reproducibility issues and positively impact population health. Timely implementation of solution-focused approaches is critical as new data sources and methods take root in public-health research, including urban public health and digital epidemiology. This commentary highlights methodological and analytic approaches that can reduce research waste and improve the reproducibility and replicability of big data research in public health. The recommendations described in this commentary, including a focus on practices, publication norms, and education, are neither exhaustive nor unique to big data, but, nonetheless, implementing them can broadly improve public-health research. Clearly defined and openly shared guidelines will not only improve the quality of current research practices but also initiate change at multiple levels: the individual level, the institutional level, and the international level.

Keywords: reproducibility; big data; digital epidemiology; urban public health

1. Introduction

Research comprises “creative and systematic work undertaken in order to increase the stock of knowledge” [1,2]. Research waste, or research whose results offer no social benefit [3], was characterized in a landmark series of papers in the *Lancet* in 2014 [4,5]. The underlying drivers of research waste range from methodological weaknesses in specific studies to systemic shortcomings within the broader research ecosystem, notably including a reward system that incentivises quantity over quality and incentivizes exploring new hypotheses over confirming old ones [4–8].

Published research that cannot be reproduced is wasteful due to doubts about its quality and reliability. Lack of reproducibility is a concern in all scientific research, and it is especially significant in the field of public health, where research aims to improve treatment practices and policies that have widespread implications. In this commentary, we highlight the urgency of improving norms for reproducibility and scientific integrity

in urban public health and digital epidemiology and discuss potential approaches. We first discuss some examples of big data sources and their uses in urban public health, digital epidemiology, and other fields, and consider the limitations with the use of big data. We then provide an overview of relevant solutions to address the key challenges to reproducibility and scientific integrity. Finally, we consider some of their expected outcomes, challenges, and implications.

Unreliable research findings also represent a serious challenge in public-health research. While the peer-review process is designed to ensure the quality and integrity of scientific publications, the implementation of peer review varies between journals and disciplines and does not guarantee that the data used are properly collected or employed. As a result, reproducibility remains a challenge. This is also true in the context of the emerging field of big data science. This is largely driven by the characteristics of big data, such as their volume, variety, and velocity, as well as the novelty and excitement surrounding new data science methods, lack of established reporting standards, and a nascent field that continues to change rapidly in parallel to the development of new technological and analytic innovations. Recent reports have uncovered that most research is not reproducible, with findings casting doubt on the scientific integrity of much of the current research landscape [6,9–12]. At the bottom of this reproducibility crisis lies growing pressure to publish not only novel, but more importantly, statistically significant results at an accelerated pace [13,14], increasing the use of low standards of evidence and disregarding pragmatic metrics, such as clinical or practical significance [15]. Consequently, the credibility of scientific findings is decreasing, potentially leading to cynicism or reputational damage to the research community [16,17]. Addressing the reproducibility crisis is not only one step towards restoring the public's trust in scientific research, but also a necessary foundation for future research, as well as guiding evidence-based public-health initiatives and policies [18], facilitating translation and implementation of research findings [19,20], and accelerating scientific discovery [21].

While failure to fully document the scientific steps taken in a research project is a fundamental challenge across all research, big data research is additionally burdened by the technical and computational complexities of handling and analysing large datasets. The challenge of ensuring computational capacity, including memory and processing power, to handle the data, as well as statistical and subject matter expertise accounting for data heterogeneity, can lead to reproducibility issues at a more pragmatic level. For example, large datasets derived from social media platforms require data analysis infrastructure, software, and technical skills, which are not always accessible to every research team [22,23]. Likewise, studies involving big data create new methodological challenges for researchers as the complexity for analysis and reporting increases [24]. This complexity not only requires sophisticated statistical skills but also new guidelines that define how data should be processed, shared, and communicated to guarantee reproducibility and maintain scientific integrity, while protecting private and sensitive information. Some of these challenges lie beyond the abilities and limitations of individual researchers and even institutions, requiring cultural and systematic changes to improve not only the reproducibility but also transparency and quality of big data research in public health.

Importantly, through concerted efforts and collaboration across disciplines, there are opportunities to systematically identify and address this reproducibility crisis and to specifically apply these approaches to big data research in public health. Below, we discuss methodological and analytical approaches to address the previously discussed issues, reduce waste, and improve the reproducibility and replicability of big data research in public health.

Specifically, we focus on approaches to improve reproducibility, which is distinct from replicability. While both are important with regards to research ethics, replicability is about “obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data”, whereas reproducibility refers to “obtaining consistent results using the same input data’ computational steps, methods and

code, and conditions of analysis” [25]. Though we mention “reproducibility” throughout this commentary, some of the arguments presented may apply to replicability as well. This is particularly true when it comes to transparency when reporting sampling, data collection, aggregation, inference methods, and study context; these affect both replication and reproduction [26].

2. Big Data Sources and Uses in Urban Public Health and Digital Epidemiology

Big data, as well as relevant methods and analytical approaches, have gained increasing popularity in recent years. This is reflected in the growing number of publications and research studies that have implemented big data methods across a variety of fields and sectors, such as manufacturing [27], supply-chain management [28], sports [29], education [30], and public health [31].

Public health, including urban health and epidemiological research, is a field where studies increasingly rely on big data methods, such as in the relatively new field of digital epidemiology [32]. The use of big data in public-health research is often characterized by the ‘3Vs’: variety in types of data as well as purposes; volume or amount of data; and velocity, referring to the speed at which the data are generated [33]. Because large datasets can invariably produce statistically significant findings but systematic biases are unaffected by data scale, big data studies are at greater risk of producing inaccurate results [34–37].

Big data sources that are used or could be potentially used in fields, such as urban public health and digital epidemiology, can be divided into two main categories. First, those that are collected or generated with health as a main focus and, second, those that are generated out of this scope but that can be associated with or impact public health (Figure 1) [32].

EXAMPLES OF BIG DATA SOURCES FOR URBAN PUBLIC HEALTH AND DIGITAL EPIDEMIOLOGY

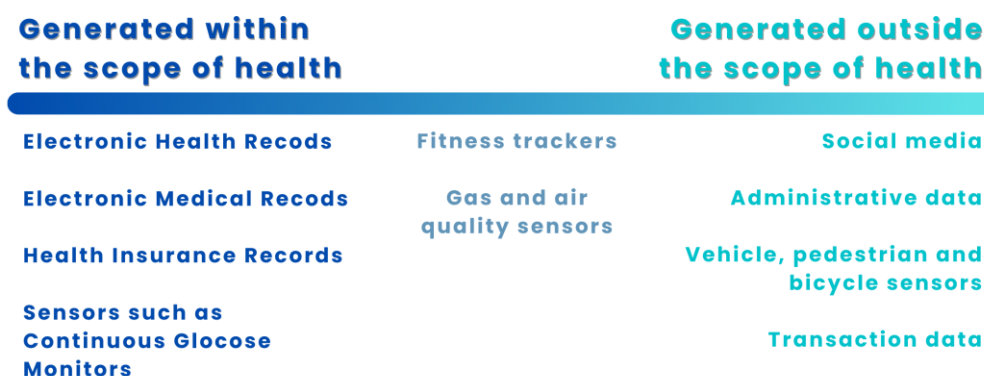


Figure 1. Data sources used in urban public health and digital epidemiology research can broadly be organized along a continuum of health orientation of the process that generated them.

Data sources generated within the context of public health include large datasets captured within health systems or government health services at the population level, such as the case of Electronic Health Records (EHRs), Electronic Medical Records (EMRs), or personal health records (PHRs) [38]. Other examples include pharmacy and insurance records, omics data, as well as data collected by sensors and devices that are part of the internet of things (IoT) and are used for health purposes, ranging from smart continuous glucose monitors (CGMs) [39] to activity and sleep trackers.

In contrast, big data sources generated outside the public-health scope are virtually unlimited and ever-growing, covering virtually all domains in society. As a result, we will focus on selected and non-conclusive examples to illustrate and exemplify the diverse sources of big data that are used or could potentially be used in urban public health and

digital epidemiology. Notably, social media have become an important source of big data used for research in different fields, including digital epidemiology. Twitter data have proven to be useful for collecting public-health information, for example, to measure mental health in different patient subgroups [40]. Examples of big data collected on Twitter that can be used in the context of public-health research are the Harvard CGA Geotweet Archive [41] or the University of Zurich Social Media Mental Health Surveillance project with their Geotweet Repository for the wider European Region [42]. Other initiatives, such as the SoBigData Research Infrastructure (RI), aim to foster reproducible and ethical research through the creation of a ‘Social Mining & Big Data Ecosystem’, allowing for the comparison, re-use, and integration of big data, methods, and services into research [22].

Cities increasingly use technological solutions, including IoT and multiple sensors, to monitor the urban environment, transitioning into Smart Cities with the objective of improving citizens’ quality of life [43,44]. Data stemming from Smart City applications have been used, for example, to predict air quality [45], analyse transportation to improve road safety [46], and have the potential to inform urban planning and policy design to build healthier and more sustainable cities [47].

Data mining techniques also allow for large datasets to be used in the context of urban public health and digital epidemiology. For example, a project using administrative data and data mining techniques in El Salvador identified anomalous spatiotemporal patterns of sexual violence and informed ways in which such analysis can be conducted in real time to allow for local law enforcement agencies and policy makers to respond appropriately [48,49]. Other large-dataset sources, such as transaction data [50], have been used to investigate the effect of sugar taxes [51] or labelling [52] on the consumption of healthy or unhealthy beverages and food products, which can eventually help model their potential impact on health outcomes.

3. Approaches to Improving Reproducibility and Scientific Integrity

Big data science has brought on new challenges, to which the scientific community needs to adapt by applying adequate ethical, methodological, and technological frameworks to cope with the increasing amount of data produced [53]. As a result, the timely adoption of approaches to address reproducibility and scientific integrity issues is imperative to ensure quality research and outcomes. A timely adoption is relevant not only for the scientific community but also for the general public that can potentially benefit from knowledge and advancements resulting from the use of big data research. This is particularly important in the context of urban public health and digital epidemiology, as the use of big data in these fields can help answer highly relevant and pressing descriptive (what is happening), predictive (what could happen), and prescriptive (why it happened) research questions [54]. A brief summary of the main points discussed in this section can be found in Figure 2. We divide our proposed solutions in this commentary into three main domains: (1) good research practice, (2) scientific communication and publication, and (3) education.

3.1. Good Research Practice

Practices, such as pre-registration of protocols, predefining research questions and hypotheses, publicly sharing data analysis plans, and communicating through reporting guidelines, can improve the quality and reliability of research and results [55,56]. For experimental studies, clear and complete reporting and documentation are essential to allow for reproduction. Observational studies can also be registered on well-established registries, such as on clinicaltrials.gov. Importantly, pre-registration does not preclude publishing exploratory results; rather, it encourages such endeavours to be explicitly described as exploratory, with defined hypotheses and expected outcomes, which is appropriate [35,37].



Figure 2. Approaches that address good research practice, scientific communication, and education are important to improve reproducibility and scientific integrity.

Lack of data access is another key challenge to reproducibility. Adoption of open-science practices, including sharing of data and code, represents a partial solution to this issue [57,58], acknowledging that not all data can be shared openly owing to privacy concerns. Similarly, transparent descriptions of data collection and analytic methods are necessary for reproduction [59]. For example, in the analysis of human mobility, which has applications in a wide range of fields, including public health and digital epidemiology [60,61], the inference of ‘meaningful’ locations [62] from mobility data has been approached with a multitude of methods, some of which lack sufficient documentation. Whereas a research project using an undocumented method to identify subject homes cannot be reproduced, a project using Chen and Poorthius’s [63] R package ‘homelocator’, which is open source and freely available, could be.

Likewise, a case could be made to collaboratively share big data within research networks and IT infrastructures. An example of a project tackling this issue in the context of public health is currently being developed by the Swiss Learning Health System (SLHS), focusing on the design and implementation of a metadata repository with the goal of developing Integrated Health Information Systems (HISs) in the Swiss context [64,65]. The implementation of such repositories and data-management systems allows for retrieval of

and access to information; nevertheless, as information systems develop, new challenges arise, particularly when it comes to infrastructure as well as legal and ethical issues, such as data privacy. Solutions are currently in development; it is likely that decentralised data architectures based on blockchain will play an important role in integrated care and health information models [66]. We briefly expand on this topic in the Anticipated Challenges section below.

The adoption of appropriate big data handling techniques and analytical methods is also important to ensure the findability, accessibility, interoperability, and reusability (FAIR) [67] of both data and research outcomes [68]. Such characteristics allow for different stakeholders to use and reuse data and research outcomes for further research, replication, or even implementation purposes.

Complete and standardised reporting of aspects discussed in this section, for instance, in Reproducibility Network Groups, allows for meta-research and meta-analyses, the detection and minimization of publication bias, and the evaluation of the adherence of researchers to guidelines focused on ensuring scientific integrity. The use of checklists by individual researchers, research groups, departments, or even institutions can motivate the implementation of good research practices as well as clear and transparent reporting, ultimately improving research integrity [69]. Such checklists can serve as training tools for younger researchers, as well as offer practice guidelines to ensure quality research.

Senior researchers and research institutions are vital when it comes to tackling these challenges as well. The adoption of principles for research conduct, such as the Hong Kong principles, can help minimise the use of questionable research practices [70]. These principles are to: (1) assess responsible research practices; (2) value complete reporting; (3) reward the practice of open science; (4) acknowledge a broad range of research activities; and (5) recognise essential other tasks, such as peer review and mentoring [71]. The promotion of these principles by mentors and institutions is a cornerstone of good research practices for younger researchers.

3.2. Scientific Communication

Scientific communication, not only between researchers but also between institutions, should be promoted. Recently, requirements for researchers to make data public or open source have grown popular among journals and major funding agencies in the US, Europe, and globally; this is an important catalyst for open science and addressing issues such as reproducibility [72].

Likewise, publication and sharing of protocols, data, code, analysis, and tools are important. This not only facilitates reproducibility but also promotes openness and transparency [73]. For example, the Journal of Memory and Language adopted a mandatory data-sharing policy in 2019. An evaluation of this policy implementation found that data sharing increased more than 50% and the strongest predictor for reproducibility was the sharing of analysis code, increasing the probability of reproducibility by 40% [57]. Such practices are also fostered by the creation and use of infrastructure, such as the aforementioned SoBigData, and reproducibility network groups, such as the Swiss Reproducibility Network, a peer-lead group that aims to improve both replicability and reproducibility [74], improve communication, collaboration, and encourage the use of rigorous research practices.

When publishing or communicating their work, researchers should also keep in mind that transparency regarding whether studies are exploratory (hypothesis forming) or confirmatory (hypothesis testing) is important to distinguish from testing newly formed hypotheses and the testing of existing ones [75]; this is particularly important for informing future research. Journal reviewers and referees should also motivate researchers to accurately report this.

Similarly, when publishing results, the quality, impact, and relevance of a publication should be valued more than scores, such as the impact factor, to avoid “publishing for numbers” [76]. This would, of course, require a shift in the priorities and views shared within the research community and may be a challenging change to effect.

Academic editors can also play an important role by avoiding practices, such as ‘cherry-picking’ publications, either because of statistical significance of results or notoriety of the authors. Instead, practical significance, topic relevance, and replication studies should be important factors to consider, as well as valuing the reporting of negative results. It is important to acknowledge, though, that scientific publication structures face an important number of challenges that hinder the implementation of these practices. Some of these points are mentioned in the Challenges section that follows.

3.3. Education

Academic institutions have the responsibility to educate researchers in an integral way, covering not only the correct implementation of methodological approaches and appropriate reporting but also how to conduct research in an ethical way.

First, competence and capacity building should be addressed explicitly through courses, workshops, and competence-building programs aimed at developing technical skills, good research practices, and adequate application of methods and analytical tools. Other activities such as journal clubs can allow researchers to exchange and become familiar with different methodologies, stay up to date with current knowledge and ongoing research, and develop critical thinking skills [77,78], while fostering a mindset for continuous growth and improvement.

Second, by incorporating practice-based education, particularly with research groups that already adhere to best practices, such as the Hong Kong principles, institutions can foster norms valuing reproducibility implicitly as an aspect of researcher education.

4. Expected Outcomes

Ideally, successful implementation of the approaches proposed in Figure 2, and the methodological and analytical approaches, such as the standardised protocols that were suggested by Simera et al. [55] and the Equator Network reporting guidelines [79], can potentially lead to a cultural shift in the research community. This, in turn, can enhance transparency and the quality of public-health research using big data by fostering interdisciplinary programs and worldwide cooperation among different health-related stakeholders, such as researchers, policy makers, clinicians, providers, and the public. Improving research quality can lead to greater value and reliability, while decreasing research waste, thus, improving the cost–value ratio and trust between stakeholders [80,81], and as previously stated, facilitating translation and implementation of research findings [18].

Just in the way replicability is fundamental in engineering to create functioning and reliable products or systems, replicability is also necessary for modelling and simulation in the fields of urban public health and digital epidemiology [82]. Simulation approaches built upon reproducible research allow for the construction of accurate prediction models with important implications for healthcare [83] and public health [84]. In the same way, reproduction and replication of results for model validation are essential [85–87].

The importance of reducing research waste and ensuring the value of health-related research is reflected in the existence of initiatives, such as the AllTrials Campaign, EQUATOR (enhancing the quality and transparency of health research), and EVBRES (evidence-based research), which promote protocol registration, full methods, and result reporting, and new studies that build on an existing evidence base [79,88–90].

Changes in editorial policies and practices can improve critical reflection on research quality by the authors. Having researchers, editors, and reviewers use guidelines [91], such as ARRIVE [92] in the case of pre-clinical animal studies or STROBE [93] for observational studies in epidemiology, can significantly improve reporting and transparency. For example, an observational cohort study analysing the effects of a change in the editorial policy of *Nature*, which introduced a checklist for manuscript preparation, demonstrated that reporting risk of bias improved substantially as a consequence [94].

A valuable outcome of adopting open science approaches that could result in improved communication, shared infrastructure, open data, and collaboration between re-

searchers and even institutions is the implementation of competitions, challenges, or even ‘hackathons’. These events are already common among other disciplines, such as computer science, the digital tech sector, and social media research, and are becoming increasingly popular in areas related to public health. Some examples include the Big Data Hackathon San Diego, where the theme for 2022 was ‘Tackling Real-world Challenges in Healthcare’ [95], and the Yale CBIT Healthcare Hackathon of 2021, which aimed to build solutions to challenges faced in healthcare [96]. In addition to tackling issues in innovative ways, hackathons and other similar open initiatives invite the public to learn about and engage with science [97] and can be powerful tools for engaging diverse stakeholders and training beyond the classroom [98].

5. Anticipated Challenges

While the implementation of the approaches discussed (Figure 2) will ideally translate to a significant reduction in research waste and improvement in scientific research through standardization and transparency, there are also substantial challenges to consider (Figure 3).

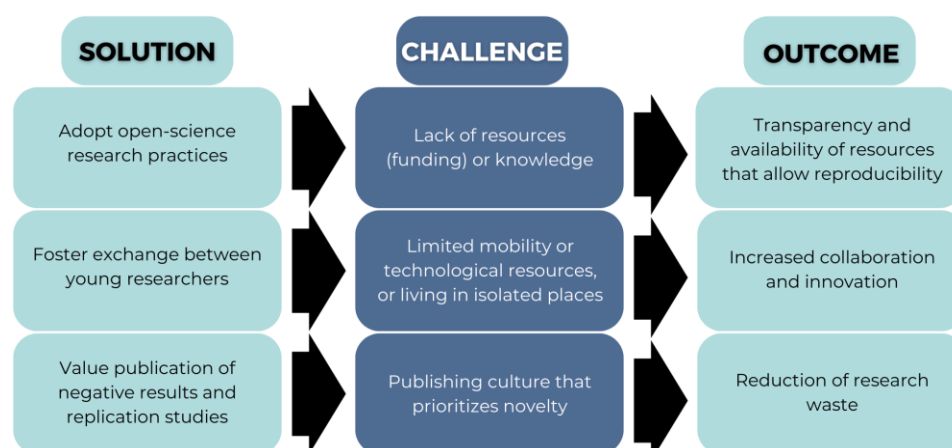


Figure 3. Examples of challenges to expect when implementing approaches aimed at improving reproducibility.

First, not all researchers have the adequate resources or opportunities to take advantage of new data that can be used to prevent, monitor, and improve population health. Early career researchers in low-resource settings may be at a particular disadvantage. Among these researchers, barriers to access and adequately using big data may not only be financial, when funding is not available, but also technical, when the knowledge and tools required are not available.

Similarly, events and activities among young researchers can facilitate technical development, networking, and knowledge acquisition, ultimately improving research quality and outcomes. Those who live in environments with limited resources, who are physically isolated, or have limited mobility may not have access to these opportunities. It might be possible to overcome some of these limitations with accessible digital solutions.

Much needed shifts in the research and publishing culture that currently enable Questionable Research Practices (QRPs), such as cherry picking (presenting favourable evidence or results while hiding unfavourable ones), p-hacking (misusing data through relentless analysis in order to obtain statistically significant results), HARKing (Hypothesizing After the Results are Known), among others [59,99,100]. To overcome these particular challenges embedded in modern day research, it is necessary to educate researchers about the scope of misconduct, create structures to avoid it from happening, and scrutinize cases in which these instances may be apparent to determine the actual motive [101].

Conventional data storing and handling strategies are not sufficient when working with big data, as these often impose additional monetary and computational costs. Some solutions are available to tackle these issues, such as cloud computing and platforms that allow end users to access shared resources over the internet [102]; Data Lakes, consisting of centralized repositories that allow for data storage and analysis [103]; and Data Mesh, a platform architecture that distributes data among several nodes [104]. Unfortunately, these solutions are not always easily accessible. Additionally, use of these platforms has given rise to important debates concerning issues, such as data governance and security [105].

The use of big data, and especially the use of personal and health information, raises privacy issues. The availability of personal and health information that results from the digital transformation represents a constant challenge when it comes to drawing a line between public and private, sensitive and non-sensitive information, and adherence to ethical research practices [106].

Ethical concerns are not limited to privacy; while big data entails the use of increasingly complex analytical methods that require expertise in order to deal with noise and uncertainty, there are several additional factors that may affect the accuracy of research results [107]. For example, when using machine learning approaches to analyse big data, methods should be cautiously chosen to avoid issues, such as undesired data-driven variable selection, algorithmic biases, and overfitting the analytic models [108]. Complexity increases the need for collaboration, which makes “team science” and other collaborative problem-solving events (such as Hackathons) increasingly popular. This leads to new requirements to adequately value and acknowledge contributorship [109].

Because statistical methods are becoming increasingly complex and the quantity of data is becoming greater, the number of scientific publications is also increasing, making it challenging for already-flawed peer-review systems to keep up by providing high-quality reviews to more and more complex research. Currently, there are mainly four expectations from peer-review processes: (i) assure quality and accuracy of research, (ii) establish a hierarchy of published work, (iii) provide fair and equal opportunities, and (iv) assure fraud-free research [110]; however, it is not certain whether current peer-review procedures achieve or are capable of delivering such expectations. Some solutions have been proposed to address these issues, such as the automation of peer-review processes [111] and the implementation of open review guidelines [112–114].

6. Conclusions

Big data research presents a unique opportunity for a cultural shift in the way public-health research is conducted today. At the same time, big data use will only result in a beneficial impact to the field if used adequately, taking the appropriate measures so that their full potential can be harnessed. The inherent complexity in working with large data quantities requires a clear and transparent framework at multiple levels, ranging from protocols and methods used by individual scientists to institution’s guiding dogma, research, and publishing practices.

The solutions summarized in this commentary are aimed at enhancing results, reproducibility, and scientific integrity; however, we acknowledge that these solutions are not exhaustive and there may be many other promising approaches to improve the integrity of big data research as it applies to public health. The solutions described in this commentary are in line with “a manifesto for reproducible science” published in the *Nature Human Behavior* journal [101]. Importantly, reproducibility is only of value if the findings are expected to have an important impact on science, health, and society. Reproducibility of results is highly relevant for funding agencies and governments, who often recognize the importance of research projects with well-structured study designs, defined data-processing steps, and transparent analysis plans (e.g., statistical analysis plans) [115,116]. For imaging data, such as radiologic images, analysis pipelines have been shown to be suitable to structure the analysis pathway [117]. This is specifically important for big data analysis where interdisciplinarity and collaboration become increasingly important. The development

and use of statistical and reporting guidelines support researchers in making their projects more reproducible [118].

Transparency in all the study-design steps (i.e., from hypothesis generation to availability of collected data and code) is specifically relevant for public health and epidemiological research in order to encourage funding agencies, the public, and other researchers and relevant stakeholders to trust research results [119]. Similarly, as globalization and digitalization increase the diffusion of infectious diseases [120] and behavioural risks [121], research practices that foster reproducible results are imperative to implement and diffuse interventions more swiftly.

We believe that these recommendations outlined in this commentary are not unique to big data and that the entire research community could benefit from the use of these approaches [122–127]. However, what has been detailed in this commentary is specifically pertinent for big data, as an increase in the volume and complexity of data produced requires more structure and consequent data handling to avoid research waste. With clearly defined and openly shared guidelines, we may strengthen the quality of current research and initiate a shift on multiple levels: at the individual level, the institutional level, and the international level. Some challenges are to be expected, particularly when it comes to finding the right incentives for these changes to stick, but we are confident that with the right effort, we can put scientific integrity back at the forefront of researchers' minds and, ultimately, strengthen the trust of the population in public-health research and, specifically, public-health research leveraging big data for urban public health and digital epidemiology.

The timely implementation of these solutions is highly relevant, not only to ensure the quality of research and scientific output, but also to potentially allow for the use of data sources that originated without public health in mind, spanning various fields that are relevant to urban public health and digital epidemiology. As outlined in this commentary, such data can originate from multiple sources, such as social media, mobile technologies, urban sensors, and GIS, to mention a few. As such data sources grow and become more readily available, it is important for researchers and the scientific community to be prepared to use these valuable and diverse data sources in innovative ways to advance research and practice. This would allow for the expanded use of big data to inform evidence-based decision making to positively impact public health.

Author Contributions: Conceptualization, A.C.Q.G., D.J.L., A.T.H., T.S. and O.G.; writing—original draft preparation, A.C.Q.G., D.J.L., A.T.H. and T.S.; writing—review and editing, A.C.Q.G., D.J.L., M.S., S.E., S.J.M., J.A.N., M.F. and O.G.; supervision, O.G. All authors have read and agreed to the published version of the manuscript.

Funding: Swiss School of Public Health (SSPH+) to O.G. This commentary is an outcome of an SSPH+ PhD course on Big Data in Public Health (website: <https://ssphplus.ch/en/graduate-campus/en/graduate-campus/course-program/> (accessed on 10 October 2022)).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful for the support of the Swiss School of Public Health (SSPH+) and, in particular, to all lecturers and participants of the class of 2021 Big Data in Public Health course. We would also like to extend our gratitude to the reviewers for their excellent feedback and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. OECD. *OECD Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*; OECD: Paris, France, 2015.
2. Grainger, M.J.; Bolam, F.C.; Stewart, G.B.; Nilsen, E.B. Evidence Synthesis for Tackling Research Waste. *Nat. Ecol. Evol.* **2020**, *4*, 495–497. [[CrossRef](#)] [[PubMed](#)]

3. Glasziou, P.; Chalmers, I. Research Waste Is Still a Scandal—An Essay by Paul Glasziou and Iain Chalmers. *BMJ* **2018**, *363*, k4645. [[CrossRef](#)]
4. Ioannidis, J.P.A.; Greenland, S.; Hlatky, M.A.; Khoury, M.J.; Macleod, M.R.; Moher, D.; Schulz, K.F.; Tibshirani, R. Increasing Value and Reducing Waste in Research Design, Conduct, and Analysis. *Lancet* **2014**, *383*, 166–175. [[CrossRef](#)] [[PubMed](#)]
5. Salman, R.A.-S.; Beller, E.; Kagan, J.; Hemminki, E.; Phillips, R.S.; Savulescu, J.; Macleod, M.; Wisely, J.; Chalmers, I. Increasing Value and Reducing Waste in Biomedical Research Regulation and Management. *Lancet* **2014**, *383*, 176–185. [[CrossRef](#)] [[PubMed](#)]
6. Begley, C.G.; Ellis, L.M. Raise Standards for Preclinical Cancer Research. *Nature* **2012**, *483*, 531–533. [[CrossRef](#)]
7. Nosek, B.A.; Hardwicke, T.E.; Moshontz, H.; Allard, A.; Corker, K.S.; Dreber, A.; Fidler, F.; Hilgard, J.; Kline Struhl, M.; Nuijten, M.B.; et al. Replicability, Robustness, and Reproducibility in Psychological Science. *Annu. Rev. Psychol.* **2022**, *73*, 719–748. [[CrossRef](#)]
8. Mesquida, C.; Murphy, J.; Lakens, D.; Warne, J. Replication Concerns in Sports and Exercise Science: A Narrative Review of Selected Methodological Issues in the Field. *R. Soc. Open Sci.* **2022**, *9*, 220946. [[CrossRef](#)]
9. Ioannidis, J.P.A. Why Most Published Research Findings Are False. *PLoS Med.* **2005**, *2*, e124. [[CrossRef](#)]
10. Raff, E. A Step Toward Quantifying Independently Reproducible Machine Learning Research. In Proceedings of the Advances in Neural Information Processing Systems; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32.
11. Hudson, R. Should We Strive to Make Science Bias-Free? A Philosophical Assessment of the Reproducibility Crisis. *J. Gen. Philos. Sci.* **2021**, *52*, 389–405. [[CrossRef](#)]
12. Block, J.A. The Reproducibility Crisis and Statistical Review of Clinical and Translational Studies. *Osteoarthr. Cartil.* **2021**, *29*, 937–938. [[CrossRef](#)]
13. Baker, M. 1500 Scientists Lift the Lid on Reproducibility. *Nature* **2016**, *533*, 452–454. [[CrossRef](#)]
14. Munafò, M.R.; Chambers, C.D.; Collins, A.M.; Fortunato, L.; Macleod, M.R. Research Culture and Reproducibility. *Trends Cogn. Sci.* **2020**, *24*, 91–93. [[CrossRef](#)]
15. Benjamin, D.J.; Berger, J.O.; Johannesson, M.; Nosek, B.A.; Wagenmakers, E.-J.; Berk, R.; Bollen, K.A.; Brembs, B.; Brown, L.; Camerer, C.; et al. Redefine Statistical Significance. *Nat. Hum. Behav.* **2018**, *2*, 6–10. [[CrossRef](#)]
16. Skelly, A.C. Credibility Matters: Mind the Gap. *Evid. Based Spine Care J.* **2014**, *5*, 2–5. [[CrossRef](#)]
17. Romero, F. Philosophy of Science and the Replicability Crisis. *Philos. Compass* **2019**, *14*, e12633. [[CrossRef](#)]
18. Perry, C.J.; Lawrence, A.J. Hurdles in Basic Science Translation. *Front. Pharmacol.* **2017**, *8*, 478. [[CrossRef](#)]
19. Reynolds, P.S. Between Two Stools: Preclinical Research, Reproducibility, and Statistical Design of Experiments. *BMC Res. Notes* **2022**, *15*, 73. [[CrossRef](#)]
20. Haymond, S.; Master, S.R. How Can We Ensure Reproducibility and Clinical Translation of Machine Learning Applications in Laboratory Medicine? *Clin. Chem.* **2022**, *68*, 392–395. [[CrossRef](#)]
21. Grant, S.; Wendt, K.E.; Leadbeater, B.J.; Supplee, L.H.; Mayo-Wilson, E.; Gardner, F.; Bradshaw, C.P. Transparent, Open, and Reproducible Prevention Science. *Prev. Sci.* **2022**, *23*, 701–722. [[CrossRef](#)]
22. Giannotti, F.; Trasarti, R.; Bontcheva, K.; Grossi, V. SoBigData: Social Mining & Big Data Ecosystem. In Companion Proceedings of the Web Conference 2018, Lyon, France, 23–27 April 2018; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2018; pp. 437–438.
23. Trilling, D.; Jonkman, J.G.F. Scaling up Content Analysis. *Commun. Methods Meas.* **2018**, *12*, 158–174. [[CrossRef](#)]
24. Olteanu, A.; Castillo, C.; Diaz, F.; Kiciman, E. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Front. Big Data* **2019**, *2*, 13. [[CrossRef](#)] [[PubMed](#)]
25. National Academies of Sciences, Engineering, and Medicine, Policy and Global Affairs; Committee on Science, Engineering, Medicine, and Public Policy; Board on Research Data and Information; Division on Engineering and Physical Sciences; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Analytics; Division on Earth and Life Studies; Nuclear and Radiation Studies Board; Division of Behavioral and Social Sciences and Education; Committee on National Statistics; et al. *Understanding Reproducibility and Replicability*; National Academies Press: Washington, DC, USA, 2019.
26. Hensel, P.G. Reproducibility and Replicability Crisis: How Management Compares to Psychology and Economics—A Systematic Review of Literature. *Eur. Manag. J.* **2021**, *39*, 577–594. [[CrossRef](#)]
27. Bertonce, T.; Meško, M.; Bach, M.P. Big Data for Smart Factories: A Bibliometric Analysis. In Proceedings of the 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 20–24 May 2019; pp. 1261–1265.
28. Mishra, D.; Gunasekaran, A.; Papadopoulos, T.; Childe, S.J. Big Data and Supply Chain Management: A Review and Bibliometric Analysis. *Ann. Oper. Res.* **2018**, *270*, 313–336. [[CrossRef](#)]
29. Šuštaršič, A.; Videmšek, M.; Karpljuk, D.; Miloloža, I.; Meško, M. Big Data in Sports: A Bibliometric and Topic Study. *Bus. Syst. Res. Int. J. Soc. Adv. Innov. Res. Econ.* **2022**, *13*, 19–34. [[CrossRef](#)]
30. Marín-Marín, J.-A.; López-Belmonte, J.; Fernández-Campoy, J.-M.; Romero-Rodríguez, J.-M. Big Data in Education. *A Bibliometric Review. Soc. Sci.* **2019**, *8*, 223. [[CrossRef](#)]
31. Galetsi, P.; Katsaliaki, K. Big Data Analytics in Health: An Overview and Bibliometric Study of Research Activity. *Health Inf. Libr. J.* **2020**, *37*, 5–25. [[CrossRef](#)]
32. Salathé, M. Digital Epidemiology: What Is It, and Where Is It Going? *Life Sci. Soc. Policy* **2018**, *14*, 1. [[CrossRef](#)]
33. Mooney, S.J.; Westreich, D.J.; El-Sayed, A.M. Epidemiology in the Era of Big Data. *Epidemiology* **2015**, *26*, 390–394. [[CrossRef](#)]

34. Smith, D.L. Health Care Disparities for Persons with Limited English Proficiency: Relationships from the 2006 Medical Expenditure Panel Survey (MEPS). *J. Health Disparit. Res. Pract.* **2010**, *3*, 11.
35. Glymour, M.M.; Osypuk, T.L.; Rehkopf, D.H. Invited Commentary: Off-Roading with Social Epidemiology—Exploration, Causation, Translation. *Am. J. Epidemiol.* **2013**, *178*, 858–863. [[CrossRef](#)]
36. Lin, M.; Lucas, H.C.; Shmueli, G. Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem. *Inf. Syst. Res.* **2013**, *24*, 906–917. [[CrossRef](#)]
37. Fan, J.; Han, F.; Liu, H. Challenges of Big Data Analysis. Available online: <https://academic.oup.com/nsr/article/1/2/293/1397586> (accessed on 8 August 2022).
38. Dash, S.; Shakyawar, S.K.; Sharma, M.; Kaushik, S. Big Data in Healthcare: Management, Analysis and Future Prospects. *J. Big Data* **2019**, *6*, 54. [[CrossRef](#)]
39. Rumbold, J.M.M.; O’Kane, M.; Philip, N.; Pierscionek, B.K. Big Data and Diabetes: The Applications of Big Data for Diabetes Care Now and in the Future. *Diabet. Med.* **2020**, *37*, 187–193. [[CrossRef](#)]
40. Hswen, Y.; Gopaluni, A.; Brownstein, J.S.; Hawkins, J.B. Using Twitter to Detect Psychological Characteristics of Self-Identified Persons with Autism Spectrum Disorder: A Feasibility Study. *JMIR Mhealth Uhealth* **2019**, *7*, e12264. [[CrossRef](#)]
41. Lewis, B.; Kakkar, D. *Harvard CGA Geotweet Archive v2.0*; Harvard University: Cambridge, MA, USA, 2022.
42. University of Zurich; Università della Svizzera italiana; Swiss School of Public Health. Emotions in Geo-Referenced Tweets in the European Region 2015–2018. Available online: https://givauzh.shinyapps.io/tweets_app/ (accessed on 12 January 2023).
43. Pivar, J. Conceptual Model of Big Data Technologies Adoption in Smart Cities of the European Union. *Entren. Enterp. Res. Innov.* **2020**, *6*, 572–585.
44. Smart Cities. Available online: https://ec.europa.eu/info/eu-regional-and-urban-development/topics/cities-and-urban-development/city-initiatives/smart-cities_en (accessed on 7 December 2022).
45. Iskandaryan, D.; Ramos, F.; Trilles, S. Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review. *Appl. Sci.* **2020**, *10*, 2401. [[CrossRef](#)]
46. Fantin Irudaya Raj, E.; Appadurai, M. Internet of Things-Based Smart Transportation System for Smart Cities. In *Intelligent Systems for Social Good: Theory and Practice; Advanced Technologies and Societal Change*; Mukherjee, S., Muppalaneni, N.B., Bhattacharya, S., Pradhan, A.K., Eds.; Springer Nature: Singapore, 2022; pp. 39–50, ISBN 978-981-19077-0-8.
47. Tella, A.; Balogun, A.-L. GIS-Based Air Quality Modelling: Spatial Prediction of PM10 for Selangor State, Malaysia Using Machine Learning Algorithms. *Env. Sci. Pollut. Res.* **2022**, *29*, 86109–86125. [[CrossRef](#)]
48. Gender Equality and Big Data: Making Gender Data Visible. Available online: <https://www.unwomen.org/en/digital-library/publications/2018/1/gender-equality-and-big-data> (accessed on 9 December 2022).
49. De-Arteaga, M.; Dubrawski, A. Discovery of Complex Anomalous Patterns of Sexual Violence in El Salvador. *arXiv* **2017**, arXiv:1711.06538v1. [[CrossRef](#)]
50. Hersh, J.; Harding, M. Big Data in Economics. *IZA World Labor* **2018**. [[CrossRef](#)]
51. Lu, X.H.; Mamiya, H.; Vybihal, J.; Ma, Y.; Buckeridge, D.L. Application of Machine Learning and Grocery Transaction Data to Forecast Effectiveness of Beverage Taxation. *Stud. Health Technol. Inform.* **2019**, *264*, 248–252. [[CrossRef](#)]
52. Petimar, J.; Zhang, F.; Cleveland, L.P.; Simon, D.; Gortmaker, S.L.; Polacsek, M.; Bleich, S.N.; Rimm, E.B.; Roberto, C.A.; Block, J.P. Estimating the Effect of Calorie Menu Labeling on Calories Purchased in a Large Restaurant Franchise in the Southern United States: Quasi-Experimental Study. *BMJ* **2019**, *367*, 15837. [[CrossRef](#)] [[PubMed](#)]
53. McCoach, D.B.; Dineen, J.N.; Chafouleas, S.M.; Briesch, A. Reproducibility in the Era of Big Data. In *Big Data Meets Survey Science*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2020; pp. 625–655, ISBN 978-1-118-97635-7.
54. Big Data and Development: An Overview. Available online: <https://datapopalliance.org/publications/big-data-and-development-an-overview/> (accessed on 9 December 2022).
55. Simera, I.; Moher, D.; Hirst, A.; Hoey, J.; Schulz, K.F.; Altman, D.G. Transparent and Accurate Reporting Increases Reliability, Utility, and Impact of Your Research: Reporting Guidelines and the EQUATOR Network. *BMC Med.* **2010**, *8*, 24. [[CrossRef](#)]
56. Equator Network. Enhancing the QUALity and Transparency of Health Research. Available online: <https://www.equator-network.org/> (accessed on 17 October 2021).
57. Laurinavichyute, A.; Yadav, H.; Vasisht, S. Share the Code, Not Just the Data: A Case Study of the Reproducibility of Articles Published in the Journal of Memory and Language under the Open Data Policy. *J. Mem. Lang.* **2022**, *125*, 104332. [[CrossRef](#)]
58. Stewart, S.L.K.; Pennington, C.R.; da Silva, G.R.; Ballou, N.; Butler, J.; Dienes, Z.; Jay, C.; Rossit, S.; Samara, A.U.K. Reproducibility Network (UKRN) Local Network Leads Reforms to Improve Reproducibility and Quality Must Be Coordinated across the Research Ecosystem: The View from the UKRN Local Network Leads. *BMC Res. Notes* **2022**, *15*, 58. [[CrossRef](#)] [[PubMed](#)]
59. Wright, P.M. Ensuring Research Integrity: An Editor’s Perspective. *J. Manag.* **2016**, *42*, 1037–1043. [[CrossRef](#)]
60. Brdar, S.; Gavrić, K.; Čulibrk, D.; Crnojević, V. Unveiling Spatial Epidemiology of HIV with Mobile Phone Data. *Sci. Rep.* **2016**, *6*, 19342. [[CrossRef](#)]
61. Fillekes, M.P.; Giannouli, E.; Kim, E.-K.; Zijlstra, W.; Weibel, R. Towards a Comprehensive Set of GPS-Based Indicators Reflecting the Multidimensional Nature of Daily Mobility for Applications in Health and Aging Research. *Int. J. Health Geogr.* **2019**, *18*, 17. [[CrossRef](#)]
62. Ahas, R.; Silm, S.; Järv, O.; Saluveer, E.; Tiru, M. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *J. Urban Technol.* **2010**, *17*, 3–27. [[CrossRef](#)]

63. Chen, Q.; Poorthuis, A. Identifying Home Locations in Human Mobility Data: An Open-Source R Package for Comparison and Reproducibility. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 1425–1448. [CrossRef]
64. Schussel  Fillietaz, S.; Berchtold, P.; Kohler, D.; Peytremann-Bridevaux, I. Integrated Care in Switzerland: Results from the First Nationwide Survey. *Health Policy* **2018**, *122*, 568–576. [CrossRef]
65. Maalouf, E.; Santo, A.D.; Cotofrei, P.; Stoffel, K. Design Principles of a Central Metadata Repository as a Key Element of an Integrated Health Information System. SLSH: Lucerne, Switzerland, 2020; 44p.
66. Tapscott, D.; Tapscott, A. What Blockchain Could Mean For Your Health Data. Harvard Business Review. 12 June 2020. Available online: <https://hbr.org/2020/06/what-blockchain-could-mean-for-your-health-data> (accessed on 17 November 2022).
67. FAIR Principles. Available online: <https://www.go-fair.org/fair-principles/> (accessed on 7 December 2022).
68. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]
69. Kretser, A.; Murphy, D.; Bertuzzi, S.; Abraham, T.; Allison, D.B.; Boor, K.J.; Dwyer, J.; Grantham, A.; Harris, L.J.; Hollander, R.; et al. Scientific Integrity Principles and Best Practices: Recommendations from a Scientific Integrity Consortium. *Sci. Eng. Ethics* **2019**, *25*, 327–355. [CrossRef]
70. Moher, D.; Bouter, L.; Kleinert, S.; Glasziou, P.; Sham, M.H.; Barbour, V.; Coriat, A.-M.; Foeger, N.; Dirnagl, U. The Hong Kong Principles for Assessing Researchers: Fostering Research Integrity. *PLoS Biol.* **2020**, *18*, e3000737. [CrossRef]
71. Bouter, L. Hong Kong Principles. Available online: <https://wcrif.org/guidance/hong-kong-principles> (accessed on 8 August 2022).
72. Bafeta, A.; Bobe, J.; Clucas, J.; Gonsalves, P.P.; Gruson-Daniel, C.; Hudson, K.L.; Klein, A.; Krishnakumar, A.; McCollister-Slipp, A.; Lindner, A.B.; et al. Ten Simple Rules for Open Human Health Research. *PLoS Comput. Biol.* **2020**, *16*, e1007846. [CrossRef]
73. Beam, A.L.; Manrai, A.K.; Ghassemi, M. Challenges to the Reproducibility of Machine Learning Models in Health Care. *JAMA* **2020**, *323*, 305–306. [CrossRef]
74. SwissRN. Available online: <http://www.swissrn.org/> (accessed on 8 August 2022).
75. Nilsen, E.B.; Bowler, D.E.; Linnell, J.D.C. Exploratory and Confirmatory Research in the Open Science Era. *J. Appl. Ecol.* **2020**, *57*, 842–847. [CrossRef]
76. Curry, S. Let’s Move beyond the Rhetoric: It’s Time to Change How We Judge Research. *Nature* **2018**, *554*, 147. [CrossRef]
77. Honey, C.P.; Baker, J.A. Exploring the Impact of Journal Clubs: A Systematic Review. *Nurse Educ. Today* **2011**, *31*, 825–831. [CrossRef]
78. Lucia, V.C.; Swanberg, S.M. Utilizing Journal Club to Facilitate Critical Thinking in Pre-Clinical Medical Students. *Int. J. Med. Educ.* **2018**, *9*, 7–8. [CrossRef]
79. EQUATOR. Available online: <https://www.equator-network.org/about-us/> (accessed on 17 November 2021).
80. Concannon, T.W.; Fuster, M.; Saunders, T.; Patel, K.; Wong, J.B.; Leslie, L.K.; Lau, J. A Systematic Review of Stakeholder Engagement in Comparative Effectiveness and Patient-Centered Outcomes Research. *J. Gen. Intern. Med.* **2014**, *29*, 1692–1701. [CrossRef]
81. Meehan, J.; Menzies, L.; Michaelides, R. The Long Shadow of Public Policy; Barriers to a Value-Based Approach in Healthcare Procurement. *J. Purch. Supply Manag.* **2017**, *23*, 229–241. [CrossRef]
82. Mulugeta, L.; Drach, A.; Erdemir, A.; Hunt, C.A.; Horner, M.; Ku, J.P.; Myers, J.G., Jr.; Vadigepalli, R.; Lytton, W.W. Credibility, Replicability, and Reproducibility in Simulation for Biomedicine and Clinical Applications in Neuroscience. *Front. Neuroinform.* **2018**, *12*, 18. [CrossRef] [PubMed]
83. Kuo, Y.-H.; Leung, J.; Tsoi, K.; Meng, H.; Graham, C. Embracing Big Data for Simulation Modelling of Emergency Department Processes and Activities. In Proceedings of the 2015 IEEE International Congress on Big Data, Santa Clara, CA, USA, 29 October–1 November 2015.
84. Belbasis, L.; Panagiotou, O.A. Reproducibility of Prediction Models in Health Services Research. *BMC Res. Notes* **2022**, *15*, 204. [CrossRef] [PubMed]
85. Schwander, B.; Nuijten, M.; Evers, S.; Hiligsmann, M. Replication of Published Health Economic Obesity Models: Assessment of Facilitators, Hurdles and Reproduction Success. *Pharmacoeconomics* **2021**, *39*, 433–446. [CrossRef] [PubMed]
86. Mahmood, I.; Arabnejad, H.; Suleimenova, D.; Sassoon, I.; Marshan, A.; Serrano-Rico, A.; Louvieris, P.; Anagnostou, A.; Taylor, S.J.E.; Bell, D.; et al. FACS: A Geospatial Agent-Based Simulator for Analysing COVID-19 Spread and Public Health Measures on Local Regions. *J. Simul.* **2022**, *16*, 355–373. [CrossRef]
87. Rand, W.; Wilensky, U. *Verification and Validation through Replication: A Case Study Using Axelrod and Hammond’s Ethnocentrism Model*; North American Association for Computational Social and Organizational Science (NAACSOS): Pittsburgh, PA, USA, 2006.
88. AllTrials All Trials Registered. All Results Reported. AllTrials 2014. Available online: <http://www.alltrials.net> (accessed on 17 November 2021).
89. Chinnery, F.; Dunham, K.M.; van der Linden, B.; Westmore, M.; Whitlock, E. Ensuring Value in Health-Related Research. *Lancet* **2018**, *391*, 836–837. [CrossRef]
90. EVBRES. Available online: <https://evbres.eu/> (accessed on 17 November 2021).
91. Shanahan, D.R.; Lopes de Sousa, I.; Marshall, D.M. Simple Decision-Tree Tool to Facilitate Author Identification of Reporting Guidelines during Submission: A before–after Study. *Res. Integr. Peer Rev.* **2017**, *2*, 20. [CrossRef]
92. Percie du Sert, N.; Hurst, V.; Ahluwalia, A.; Alam, S.; Avey, M.T.; Baker, M.; Browne, W.J.; Clark, A.; Cuthill, I.C.; Dirnagl, U.; et al. The ARRIVE Guidelines 2.0: Updated Guidelines for Reporting Animal Research. *J. Cereb. Blood Flow Metab.* **2020**, *40*, 1769–1777. [CrossRef]

93. STROBE. Available online: <https://www.strobe-statement.org/> (accessed on 14 November 2022).
94. Group, T.N.C. Did a Change in Nature Journals' Editorial Policy for Life Sciences Research Improve Reporting? *BMJ Open Sci.* **2019**, *3*, e000035. [[CrossRef](#)]
95. Big Data Hackathon for San Diego 2022. Available online: <https://bigdataforsandiego.github.io/> (accessed on 8 August 2022).
96. Yale CBIT Healthcare Hackathon. Available online: <https://yale-hack-health.devpost.com/> (accessed on 8 August 2022).
97. Ramachandran, R.; Bugbee, K.; Murphy, K. From Open Data to Open Science. *Earth Space Sci.* **2021**, *8*, e2020EA001562. [[CrossRef](#)]
98. Wilson, J.; Bender, K.; DeChants, J. Beyond the Classroom: The Impact of a University-Based Civic Hackathon Addressing Homelessness. *J. Soc. Work Educ.* **2019**, *55*, 736–749. [[CrossRef](#)]
99. van Dalen, H.P.; Henkens, K. Intended and Unintended Consequences of a Publish-or-perish Culture: A Worldwide Survey. *J. Am. Soc. Inf. Sci. Technol.* **2012**, *63*, 1282–1293. [[CrossRef](#)]
100. Andrade, C. HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. *J. Clin. Psychiatr.* **2021**, *82*, 20f13804. [[CrossRef](#)]
101. Munafò, M.R.; Nosek, B.A.; Bishop, D.V.M.; Button, K.S.; Chambers, C.D.; Percie du Sert, N.; Simonsohn, U.; Wagenmakers, E.-J.; Ware, J.J.; Ioannidis, J.P.A. A Manifesto for Reproducible Science. *Nat. Hum. Behav.* **2017**, *1*, 1–9. [[CrossRef](#)]
102. Qi, W.; Sun, M.; Hosseini, S.R.A. Facilitating Big-Data Management in Modern Business and Organizations Using Cloud Computing: A Comprehensive Study. *J. Manag. Organ.* **2022**, 1–27. [[CrossRef](#)]
103. Thomas, K.; Praseetha, N. Data Lake: A Centralized Repository. *Int. Res. J. Eng. Technol.* **2020**, *7*, 2978–2981.
104. Machado, I.A.; Costa, C.; Santos, M.Y. Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures. *Procedia Comput. Sci.* **2022**, *196*, 263–271. [[CrossRef](#)]
105. Fadler, M.; Legner, C. Data Ownership Revisited: Clarifying Data Accountabilities in Times of Big Data and Analytics. *J. Bus. Anal.* **2022**, *5*, 123–139. [[CrossRef](#)]
106. Mostert, M.; Bredenoord, A.L.; Biesaart, M.C.I.H.; van Delden, J.J.M. Big Data in Medical Research and EU Data Protection Law: Challenges to the Consent or Anonymise Approach. *Eur. J. Hum. Genet.* **2016**, *24*, 956–960. [[CrossRef](#)] [[PubMed](#)]
107. Hariri, R.H.; Fredericks, E.M.; Bowers, K.M. Uncertainty in Big Data Analytics: Survey, Opportunities, and Challenges. *J. Big Data* **2019**, *6*, 44. [[CrossRef](#)]
108. Mooney, S.J.; Keil, A.P.; Westreich, D.J. Thirteen Questions About Using Machine Learning in Causal Research (You Won't Believe the Answer to Number 10!). *Am. J. Epidemiol.* **2021**, *190*, 1476–1482. [[CrossRef](#)] [[PubMed](#)]
109. Bennett, L.M.; Gadlin, H. Collaboration and Team Science: From Theory to Practice. *J. Investig. Med.* **2012**, *60*, 768–775. [[CrossRef](#)] [[PubMed](#)]
110. Horbach, S.P.J.M.; Halffman, W. The Changing Forms and Expectations of Peer Review. *Res. Integr. Peer Rev.* **2018**, *3*, 8. [[CrossRef](#)] [[PubMed](#)]
111. Yuan, W.; Liu, P.; Neubig, G. Can We Automate Scientific Reviewing? *J. Artif. Intell. Res.* **2022**, *75*, 171–212. [[CrossRef](#)]
112. Allen, C.; Mehler, D.M.A. Open Science Challenges, Benefits and Tips in Early Career and Beyond. *PLoS Biol.* **2019**, *17*, e3000246. [[CrossRef](#)]
113. Mirowski, P. The Future(s) of Open Science. *Soc. Stud. Sci.* **2018**, *48*, 171–203. [[CrossRef](#)]
114. Ross-Hellauer, T.; Görögh, E. Guidelines for Open Peer Review Implementation. *Res. Integr. Peer Rev.* **2019**, *4*, 4. [[CrossRef](#)]
115. Hume, K.M.; Giladi, A.M.; Chung, K.C. Factors Impacting Successfully Competing for Research Funding: An Analysis of Applications Submitted to The Plastic Surgery Foundation. *Plast. Reconstr. Surg.* **2014**, *134*, 59. [[CrossRef](#)]
116. Bloemers, M.; Montesanti, A. The FAIR Funding Model: Providing a Framework for Research Funders to Drive the Transition toward FAIR Data Management and Stewardship Practices. *Data Intell.* **2020**, *2*, 171–180. [[CrossRef](#)]
117. Gorgolewski, K.; Alfaro-Almagro, F.; Auer, T.; Bellec, P.; Capotà, M.; Chakravarty, M. BIDS Apps: Improving Ease of Use, Accessibility, and Reproducibility of Neuroimaging Data Analysis Methods. *PLoS Comput. Biol.* **2017**, *13*, e1005209. [[CrossRef](#)]
118. Agarwal, R.; Chertow, G.M.; Mehta, R.L. Strategies for Successful Patient Oriented Research: Why Did I (Not) Get Funded? *Clin. J. Am. Soc. Nephrol.* **2006**, *1*, 340–343. [[CrossRef](#)] [[PubMed](#)]
119. Harper, S. Future for Observational Epidemiology: Clarity, Credibility, Transparency. *Am. J. Epidemiol.* **2019**, *188*, 840–845. [[CrossRef](#)] [[PubMed](#)]
120. Antràs, P.; Redding, S.J.; Rossi-Hansberg, E. *Globalization and Pandemics*; National Bureau of Economic Research: Cambridge, MA, USA, 2020.
121. Ebrahim, S.; Garcia, J.; Sujudi, A.; Atrash, H. Globalization of Behavioral Risks Needs Faster Diffusion of Interventions. *Prev. Chron. Dis.* **2007**, *4*, A32.
122. Gilmore, R.O.; Diaz, M.T.; Wyble, B.A.; Yarkoni, T. Progress toward Openness, Transparency, and Reproducibility in Cognitive Neuroscience. *Ann. N. Y. Acad. Sci.* **2017**, *1396*, 5–18. [[CrossRef](#)] [[PubMed](#)]
123. Brunson, C.; Comber, A. Opening Practice: Supporting Reproducibility and Critical Spatial Data Science. *J. Geogr. Syst.* **2021**, *23*, 477–496. [[CrossRef](#)]
124. Catalá-López, F.; Caulley, L.; Ridao, M.; Hutton, B.; Husereau, D.; Drummond, M.F.; Alonso-Arroyo, A.; Pardo-Fernández, M.; Bernal-Delgado, E.; Meneu, R.; et al. Reproducible Research Practices, Openness and Transparency in Health Economic Evaluations: Study Protocol for a Cross-Sectional Comparative Analysis. *BMJ Open* **2020**, *10*, e034463. [[CrossRef](#)]
125. Wachholz, P.A. Transparency, openness, and reproducibility: GGA advances in alignment with good editorial practices and open science. *Geriatr. Gerontol. Aging* **2022**, *16*, 1–5. [[CrossRef](#)]

126. Girault, J.-A. Plea for a Simple But Radical Change in Scientific Publication: To Improve Openness, Reliability, and Reproducibility, Let's Deposit and Validate Our Results before Writing Articles. *eNeuro* **2022**, *9*. [[CrossRef](#)]
127. Schroeder, S.R.; Gaeta, L.; Amin, M.E.; Chow, J.; Borders, J.C. Evaluating Research Transparency and Openness in Communication Sciences and Disorders Journals. *J. Speech Lang. Hear. Res.* **2022**. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.