Review Article
Malays. j. med. biol. res.

Using Secondary Health Data in Research

Mohammed Al Omari

School of Medicine, Swansea University, Swansea SA2 8QA, Swansea, **UNITED KINGDOM**

*Email for Correspondence: 936122@swansea.ac.uk

ABSTRACT

The use of data in medical research that was originally collected for different purposes, known as secondary data, is an effective way to conduct reliable and cost-effective studies so as to progress knowledge in medicine. A number of serious practical, ethical and legal issues and concerns about this process exist, however. Ensuring a high level of data quality is imperative to produce reliable results, and researchers may face accessibility problems. Projects designed to alleviate these issues are underway, however, lowering the cost and increasing the access to secondary data even further. Although secondary data is de-identified to protect confidentiality, ethical problems of individual rights versus the benefit of society persist, leading some to call for a new 'macroethics' surrounding data use. Legislation to this end has been introduced in many countries, but issues relating to the exemptions it offers and its interpretability remain. To ensure that the use of secondary data in medical research can continue to accelerate the pace of development in medicine, a global effort involving technological and ethical standardization needs to be developed.

Keywords: Electronic health record, medical informatics, data mining, genetic exceptionalism, data protection

Manuscript Received: 04 January 2021 - Revised: 5 March 2021 - Accepted: 10 March 2021

INTRODUCTION

In medical research, the use of secondary data refers to applying data to research for purposes different to which it was initially collected (Schlegel & Ficheur, 2017). Using secondary data can save time and costs (Boslaugh, 2007; Harpe, 2009) and provide an accurate basis for study to allow for faster progress in health research to tackle health problems and enhance patients' quality of life (Johnston, 2014). However, practical, ethical and legal issues have been identified (Tripathy, 2013). The data must be of good quality and from reliable and dependable sources (Demchenko, Grosso, De Laat & Membrey, 2013; Cox, Martin, Van Staa, Garbe Siebert & Johnson, 2009), as poor-quality data produces unreliable results (Botsis, Hartvigsen, Chen & Weng, 2010). Ensuring consistency and accuracy of data enhances the generalisability and trustworthiness of the research findings (Hagger-Johnson et al, 2015; Foley, Shuttleworth & Martin, 2018), and it must be cost-effective (National Academy of Sciences, 2009; Burton, Banner, Elliot, Knoppers & Banks, 2017). Ethically, there are concerns surrounding sensitive secondary data in medical research so protecting confidentiality is essential (Sweeney, 2002; Tripathy, 2013). Issues arise from conflicts of interest between the rights of individuals and the requirements of society (Brakewood & Poldrack, 2013), and the techniques employed to de-identify data may be undesirable for research or impossible to implement in the age of increasing data connectivity (Black, 2003; Aamot, Kohl, Richter & Knaup-Gregori, 2013; Morrow, Boddy & Lamb, 2014). There are also issues with regard to the law (Lowrance, 2003). Legislation such as the Caldicott principles, the Data Protection Act (2018) and GDPR (General Data Protection Regulation) in the UK must be followed (Mourby et al., 2018), but these differ between countries and cannot provide ethical guidance (Safran, 2007; Mourby et al., 2018). This essay critically analyses these issues and the current major debates in the field to highlight areas that could be developed to make medical research easier, cheaper and more reliable when using secondary data.

PRACTICAL ASPECTS

Data Quality

The success of a research study using secondary data is dependent upon the quality of data collected, and it must be valid and accurate so as to make the results of the study useful (Pezoulas et al., 2019; Morrow, Boddy & Lamb, 2014). Using good quality data in the research process is required to produce reliable results as through the use of poor-quality data, the outcomes of the study might be misleading and insignificant (Botsis, Hartvigsen, Chen & Weng, 2010; Endriyas et al., 2019).

All research requires that set objectives are achieved and that credibility of the outcomes is assured, and this is no different for research using secondary data (Johnston, 2014). Proper planning should be conducted in the initial stage so that the data used fits the desired purpose of the study (Nass, Levit & Gostin, 2009), and the quality of the data used should be determined through robust assessment procedures (Weiskopf & Weng, 2013; Pezoulas et al., 2019). To demonstrate high quality, data must be accurate and consistent (Taleb, El Kassabi, Serhani, Dssouli & Bouhaddioui, 2016; Juddoo & George, 2018), but for research purposes, it must also demonstrate a high level of accessibility (Safran, 2007) and cost-effectiveness (Neumann, Sanders, Russell, Siegel & Ganiats, 2016). If all of these dimensions are met sufficiently when using secondary data, the trustworthiness of the overall study is improved (Strong, Lee & Wang, 1997; Nikiforova, 2019).

Accuracy

Accuracy is a primary importance while selecting and using secondary data in the research execution (Xiao et al., 2017; Endriyas et al., 2019). Poor-quality data decreases the relevance and authenticity of the outcomes and as such, the overall study becomes invalid and inappropriate (Weiskopf & Weng, 2013). Hence, the dataset selected must be checked for formatting compatibility with the research, and the researchers should familiarise themselves with the data (Vartanian, 2010). While selecting secondary data sources, it is important to make appropriate decisions so that good quality data can be embraced within the research study (van Mourik, van Duijn, Moons, Bonten & Lee, 2015). Nevertheless, there are complications associated with the use of the secondary data (Windle, 2010). As the data is directly extracted from already published sources, certain analytical techniques may be difficult (Nass, Levit & Gostin, 2009; Dolley, 2018) and there is a strong possibility of the data being inauthentic if it was initially collected using inaccurate methods (Peabody, Luck, Jain, Bertenthal & Glassman, 2004; Takahashi et al., 2018).

Distinguishing between good and bad data quality is a crucial research process and thus it is necessary to acquire data from accurate sources (Hasan & Padman, 2006). As modern day clinical data is stored in electronic health records (EHR), the risk of programming bugs and human error while entering patient data into the system is ever present (Hagger-Johnson et al, 2015; Motulsky et al., 2018), which can result in inaccurate data (Foley, Shuttleworth & Martin, 2018). Accordingly, to ensure that secondary data is rigorously accurate, clinical audits need to be implemented to enable identification of relevant sources of data (Harpe, 2009; Haley et al., 2012).

Consistency

Consistency of data can be ensured if the data collected is reliable and uniform (Demchenko, Grosso, De Laat & Membrey, 2013). It is essential to interpret the collected data clearly in order to increase the consistency and reliability of the overall research outcomes (Castle, 2003). In clinical research, there should not be any scope for misunderstanding or misinterpretation (Iezzoni, 1997), which is particularly important to recall when extracting data from coding systems like the International Classification of Diseases (ICD) (Quan et al, 2005) or EHRs (Wilkerson, Henricks, Castellani, Whitsitt & Sinard, 2015; Prada-Ramallal, Takkouche & Figueiras, 2019) which may contain unstructured secondary data which is neither useful nor relevant for use in research studies (Kahn et al., 2016). As lack of consistency results in a loss of reliability, the consistency of secondary data must be tested before it is employed in a study (Hagger-Johnson et al, 2015; Foley, Shuttleworth & Martin, 2018).

Accessibility and cost-effectiveness

While pursuing secondary data, researchers may face accessibility problems (National Academy of Sciences, 2009). Full or partial information may be missing as a result of the lack of a good network, differing data entry and storage (Huston & Naylor, 1996; Botsis, Hartvigsen, Chen & Weng, 2010; Xiao et al., 2017) or clinician and staff EHR entry standards (Weiskopf, Hripcsak, Swaminathan & Weng, 2013). As well as this, the users of secondary data were not part of the data collection process and therefore may not know exact details of its collection (Castle, 2003; Cheng & Phillips, 2014; Johnston, 2014). These issues can be partially overcome by using data for which access to the primary researchers is possible (Windle, 2010) and greater clinician knowledge of the uses of patient data to encourage greater adherence to data entry (Kaplan, 2014; Motulsky et al., 2018).

Ensuring security and accuracy of data is necessary as medical experts cannot compromise on patient safety (Berger, Mamdani, Atkins & Johnson, 2009; Endriyas et al., 2019). Therefore, denial of authorisation regarding access to data is another hindrance faced by researchers during the process as they may not have access to specific data kept by organisations or governments (Boslaugh, 2007; Rumbold, Lewis & Bardsley, 2011) or the data may be unavailable for safeguarding purposes (Lowrance, 2003).

Thus for accessing and managing data, researchers should discover the resources and users, classify the data (Vartanian, 2010), undertake auditing after assigning data owners and automatically access and review requests through an integrated expert panel (Holzer & Gall, 2011). Permission must be granted primarily from expert officials so as to ensure the proper extraction and use of secondary data in medical research (Langarizadeh, Orooji & Sheikhtaheri, 2018).

Using secondary data is one of the most effective ways to control costs in medical research and is not necessarily any less effective or fruitful than primary data (Boslaugh, 2007; Harpe, 2009). As identifying and examining reliable and authentic sources of secondary information is the stage where the accuracy and authentication of the research process can be determined (Bhaskar & Manjuladevi, 2016), if the researchers fail to do so, it is possible to make mistakes that lead to inaccurate results. Difficulties while accessing data may reduce the trustworthiness of the research so the whole data collection process may need to be repeated, significantly increasing the cost of the research (National Academy of Sciences, 2009; Weiskopf, Hripcsak, Swaminathan & Weng, 2013).

In recent years, however, research data that can be accessed online is frequently of high quality and can be statistically analysed at a lower cost than conducting primary research (Tu, Campbell, Chen, Cauch-Dudek & McAlister, 2007; Cheng & Phillips, 2014). Researchers can refer to such sources as the NHS Health Data Finder for Research, Hospital Episode Statistics, Admitted Patient Care, Adult Critical Care, Cancer Registry Data, as these are reliable and trustworthy sources (Herbert, Wijlaars, Zylbersztejn, Cromwell & Hardelid, 2017) that contain valid and credible data that can be easily accessed and used cost-effectively. There are, however, in many cases fees for access to data (Windle, 2010; Burton, Banner, Elliot, Knoppers & Banks, 2017; Wise, 2019), so researchers need to balance the costs of access to the data against its overall utility and include consideration of fees in budget proposals (National Academy of Sciences, 2009). The advent of comprehensive national and supra-national big data curation projects and standards developing organisations has begun to increase accessibility to verified secondary data and further lowered the cost of research using secondary data (Hammond, Jaffe, & Kush, 2009; Wise, 2019; Nikiforova, 2019). Clinical research networks that use a common data model such as the EMIF (EU) and SHRINE (US) are currently in the process of extending their scope and resourcing for the purpose of easing the access, storage and transmissibility of secondary data for use in medical research (Burton, Banner, Elliot, Knoppers & Banks, 2017; Martin-Sanchez, Aguiar-Pulido, Lopez-Campos, Peek & Saachi, 2017; Pezoulas et al., 2019).

Therefore, while there can be financial costs and obstacles to accessibility when using secondary data in medical research, advances in technology, particularly with regard to interconnectivity, and the growth of large-scale data collection initiatives mean that using secondary data presents a cost-effective and increasingly convenient alternative to conducting primary research.

ETHICAL ASPECTS

It is essential to follow all ethical considerations and guidelines when conducting research, not least to ensure its validity (Tripathy, 2013). In medical research, using secondary data involves ethical measures that must be adhered to (Aita & Richer, 2005) because patients' health information constitutes sensitive data (Brakewood & Poldrack, 2013). As appropriate and accurate personal and health information about patients may be required by the research (El Emam, Rodgers & Malin, 2015), this gives rise to major issues surrounding privacy and confidentiality. During data collection from patients, medical researchers might think that the respective health issues of the patients and the associated treatment procedures might be helpful for society in general so they may desire to disclose such information and use it for their clinical studies (Brakewood & Poldrack, 2013). However, this requires informed consent from the patients which should be sought before extracting data (Kalkman, Mostert, Gerlinger, van Delden, & van Thiel, 2019).

Consent

The issue of gaining consent is critical for the use of secondary data in research as it is a patient's fundamental right to understand and agree to uses of their health information that were not initially acknowledged (Morrow, Boddy & Lamb, 2014; McLennan, Shaw & Celi, 2018). Informed consent results in voluntary research participation and thus it is important to seek consent to ensure patients' agreement to the use of their data in a particular research process (Windle, 2010; Brown, Brown & Korff, 2010). Obtaining consent may allow the use of information and health data for secondary purposes including undertaking further medical research, but for this to happen, the patients must be made aware of the reasons their data may be used for these purposes (Safran et al., 2007; Graves, McLaughlin, Leung &

Powers, 2019). Failing to adhere to this or disclosing data to any third party without such consent might lead to confidentiality breaches and privacy issues (Ploug & Holm, 2015). However, burdening subjects with consent forms or other paperwork may discourage them from allowing their data to be used in research (Singleton & Wadsworth, 2006), and obtaining informed consent may not even be possible when it is not known what the future research that uses the secondary data will entail (Thomson, Bzdel, Golden-Biddle, Reay & Estabrooks, 2005; Grinyer, 2009). Additionally, systematic bias may be introduced in cases of refused consent (Porsdam Mann, Savulescu & Sahakian, 2016).

In seeking a solution to this, it has been proposed that patients have a moral duty to be part of medical research, particularly when there is minimal risk of harm to them (Lawlor & Stone, 2001; Ballantyne & Schaefer, 2017). This is increasingly possible through technology, as is asking patients to give generalised or 'broad' consent (Hudson & Collins, 2015; Kalkman et al., 2019) or 'meta-consent' (Ploug & Holm, 2015) to appropriate secondary use of their data for research purposes (Tenenbaum et al., 2017) However, use of such consent requests must ensure that the patient is aware of the difference between the taking of their data for their own therapeutic benefit and for later use in secondary research as failing to do so would compromise the doctor-patient trust relationship (Frunză & Sandu, 2017).

Due to the sensitivity of the secondary data used in medical research, encryption is likely to be employed to protect the confidentiality of patients (Benaloh, Chase, Horvitz & Lauter, 2009; Heurix & Neubauer, 2011) as well as techniques that remove identifiable information from the data such as anonymisation. In order to execute authentic medical research, patients' personal information and secondary treatment intervention details should be extracted and used appropriately (Prada-Ramallal, Takkouche & Figueiras, 2019) in a way that can satisfy the interests of both the patient and society. Any breaches of ethical principles that take place during the use of secondary data can compromise the integrity of a research study (Ross, Iguchi & Panicker, 2018). As there are various philosophical positions to take on this, examination of two approaches, utilitarianism and deontology, is important to ensure adherence to ethical principles and the maintenance of patient confidentiality (Floridi et al., 2019).

Utilitarianism vs. Deontology

The ethics surrounding patient confidentiality and secondary use of data in research can be approached from either a deontological or a utilitarian perspective (Gray & Schein, 2012). Utilitarianism is an ethical stance through which the differences between right and wrong can be determined by considering the interests of the society as a whole rather than those of individuals, especially in cases of conflict (Pieper, 2008). In medical research, adopting this approach can assure the appropriateness of the research for the benefit of all (Lambert, Soskolne, Bergum, Howell & Dossetor, 2003; Herschel & Miori, 2017). As secondary data must fit the aims and objectives of the study concerned, if it can be proven that the outcomes of the research will benefit the whole of society, according to utilitarianism the sanctity of individual consent may be surpassed (Brown, Brown & Korff, 2010; Resnik, 2015).

In opposition to this, the deontological approach is the process of interpreting the exactness of the ethics, duties and obligations associated with each step of a research study in this case, the patients' 'right to know' about uses of their data (Lambert, Soskolne, Bergum, Howell & Dossetor, 2003). A deontological approach can assure the ethical validity of the use of secondary data in research by following the norms and principles which protect the rights of individuals (Gray & Schein, 2012). One of these norms is to maintain a high level of confidentiality and seek informed consent before using secondary data in research as deontology does not focus on the consequences of the actions performed but the acts that are undertaken at each stage (Gatignon, 2019; Mazeikiene et al., 2020). Therefore, deontology suggests that the researchers do what they feel is morally correct without allowing this to be outweighed by consideration of the social wellbeing (Beck, 2019). Thus, from a deontological perspective, researchers should seek data from the patients, protect their privacy and preserve their rights even if the social interests are not fulfilled through these actions (Coughlin, 2006; Price & Cohen, 2019).

This debate is yet unresolved. Ballantyne (2019) argues from a utilitarian perspective by advocating that individual consent should not be the utmost concern for the use of secondary data in research as this sidelines the benefits that would be gained from prioritising the greater good, whereas Frunză and Sandu (2017), conversely, adopt a deontological stance by warning that diminishing the importance of informed consent risks delegitimising the use of secondary data in research. Porsdam Mann, Suvalescu and Sahakian (2016) describe a path between the two approaches, maintaining that consent should be sought in cases of risk but it is only is one consideration among many when the benefit of research is to the public good.

It has however been argued that such narrow approaches are no longer appropriate in the highly computation-driven world of big data, and instead a new ethical framework (Xafis et al., 2019) or 'macroethics' specifically for the use of data should be developed (Floridi & Taddeo, 2016; Harsh, Achaerya & Chaudhary, 2018).

LEGAL CONSIDERATIONS

Confidentiality: anonymisation and pseudonymisation

In order to avoid breaches of confidentiality and misuse of sensitive personal information, de-identification techniques such as anonymisation and pseudonymisation are applied to secondary data that is used in research (El Emam, Rodgers & Malin, 2015; Neubauer & Heurix, 2011; Verheul, Jacobs, Meijer, Hildebrandt, & de Ruiter, 2016). Anonymisation refers to techniques to erase personally identifiable information from the data (Thomson, Bzdel, Golden-Biddle, Rea & Estabrooks, 2005; Lafoz, Ramírez-Soriano & Richardson, 2018). However, medical researchers may require a certain amount of personal or demographic information about patients such as age, gender, lifestyle and health-related issues, which is not possible through anonymisation (Heurix & Neubauer, 2011). As a result, the concept of pseudonymisation has come into prominence as it replaces patients' identifiable information with codes called pseudonyms (Pommerening & Reng, 2004; Mourby et al., 2018) which retain enough personal data to serve the aims and objectives of the research (Heurix & Neubauer, 2011; Somolinos et al, 2014).

Through the initiation of the NHS Pseudonymisation Implementation Project, researchers can view instructions on how de-identification of patients can be approached and confidentiality maintained (Stalla-Bourdillon & Knight, 2016). Privacy audits should be carried out to maximise both adherence to data use standards and research innovation (Price & Cohen, 2019). Nevertheless, in some cases the data cannot be pseudonymised (Iacono, 2007), and the suitability of de-identification techniques has been cast into doubt by the success of re-identification attacks (Sweeney, 2002; Aamot, Kohl, Richter & Knaup-Gregori, 2013). It may also be impossible to assure anonymity in the context of big data (Dolley, 2018) due to the fact that large-scale data combination compromises anonymisation (Brown, Brown & Korff, 2010; Verheul, Jacobs, Meijer, Hildebrandt, & de Ruiter, 2016; Herschel & Miori, 2017).

Accordingly, it can be stated that while anonymisation and pseudonymisation are helpful for protecting anonymity and confidentiality in the use of patients' personal data for specific secondary research purposes, questions around the utility and methodologies for this remain unresolved (Grinyer, 2009; Goroff, Polonetsky & Tene, 2018).

Some of the most comprehensive codified policy on confidentiality and use of patient data can be found in the UK NHS in the form of the Caldicott principles, established in 1997 and revised in 2013 (Taylor, 2013) and 2016 (National Data Guardian for Health, 2016) .The seven principles that determine whether data can be used are justification of use, necessity for use, minimal use of data, access on need-to-know basis, awareness of responsibilities, legal compliance and sharing data where appropriate (National Data Guardian for Health, 2016). Secondary data should be implemented in line with all of these principles and legal requirements.

THE DATA PROTECTION ACT (2018) AND GDPR (GENERAL DATA PROTECTION REGULATION)

The Data Protection Act (2018) details the UK law regarding the confidentiality and privacy of the data and information of individuals, with the broader European Union legislation known as GDPR (Mourby et al., 2018). These contain provisions related to the use of secondary data for medical research, including that no third party can access data without the consent of the individual concerned. However, the legislation makes specific and broad exemptions for research if it does not cause "substantial damage or distress" (ICO, 2020) and if pseudonymisation is applied (Mészáros, 2018). Despite this, it has been argued that this creates an imbalance by favouring data controllers at the expense of subjects (Pormeister, 2017).

The legal validity and credibility of research can be maintained if the data collected from the selected participants is kept safe and secured properly (Godard, Schmidtke, Cassiman & Aymé, 2003). In the case of healthcare research studies, the personal details and health issues of patients must be obtained and protected in line with strict laws collectively known as data protection principles (Data Protection Act 2018). The main aim of the legislation is to prevent mishandling of data, promote data protection rights and ensure optimum confidentiality and anonymity of personal and health data (Francis, 2020).

By complying with these legislated norms, medical researchers are able to securely collect, exchange and transfer the personal data of patients between medical staff (Iversen, Liddell, Fear, Hotopf & Wessely, 2006). However, there are still questions surrounding the use of some data in research, one of which is regarding the fact that the provisions of neither the Data Protection Act nor GDPR apply to non-living subjects (Data Protection Act 2018). This means that data related to deceased individuals is not considered, which may trigger serious ethical problems, differing between countries, about how medical researchers should deal with the confidentiality of this data (Edwards & Harbinja, 2013).

CONCLUSION

The use of secondary data in medical research is an area containing a number of important conflicts and questions regarding its practical, ethical and legal considerations, many of which are unresolved.

In terms of the practical aspects, the quality of the secondary data used in research must be ensured, as poor-quality data leads to unreliable and inaccurate results. Problems with the accuracy and consistency of secondary data may arise from systems such as EHR, so data authenticity must be verified and ambiguities resolved. Accessibility issues such as missing information or lack of permission may hamper research, and costs for access to secondary data need to be considered. However, the increasing prominence of large-scale data collation projects and clinical research networks with standardized data storage and transmission methods means costs are falling while accessibility is increasing.

As secondary health data is of a sensitive nature, there are significant ethical concerns that must be taken into account when employing it in research. The primary conflict in this area is between the rights of the individual and the right of society as a whole. While this can be described as a matter of utilitarianism versus deontology, it may be that neither alone can provide satisfactory ethical guidance. Asking patients for broad consent may represent a compromise, but ultimately the development of a new data ethics may be necessary.

De-identifying data can help protect confidentiality, with pseudonymization having emerged as the main way to do this, but in a world characterized by big data, true data de-identification may not be possible. While legislation such as the UK Data Protection Act 2018 has been developed to regulate the use of data, allowing for research exemptions, legal instruments do not necessarily help with the ethical problems posed by the requirement for informed consent to data sharing and use of secondary data in research.

The potential for improvements to the state of medicine through further reliable research using increasingly numerous and accessible sources of secondary data cannot be underestimated, so the need to come to international agreement on the practical, ethical and legal issues examined in this essay is now as critical as ever.

REFERENCES

Aamot, H., Kohl, C. D., Richter, D., & Knaup-Gregori, P. (2013). Pseudonymization of patient identifiers for translational research. *BMC medical informatics and decision making*, *13*(1), 75.

Aita, M., & Richer, M. C. (2005).Essentials of research ethics for healthcare professionals. *Nursing & health sciences*, *7*(2), 119-125.

Ballantyne, A. (2019). Adjusting the focus: a public health ethics approach to data research. *Bioethics*, *33*(3), 357-366.

Ballantyne, A., & Schaefer, G. O. (2017). Consent and the ethical duty to participate in health data research. *Journal of medical ethics*, *44*(6), 392-396.

Beck, C. T. (2019). Secondary qualitative data analysis in the health and social sciences. Routledge.

Benaloh, J., Chase, M., Horvitz, E., & Lauter, K. (2009). Patient controlled encryption: ensuring privacy of electronic medical records. In *Proceedings of the 2009 ACM workshop on Cloud computing security* (pp. 103-114).

Berger, M. L., Mamdani, M., Atkins, D., & Johnson, M. L. (2009). Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I. *Value in Health*, *12*(8), 1044-1052.

Bhaskar, S. B., & Manjuladevi, M. (2016). Methodology for research II. Indian journal of anaesthesia, 60(9), 646–651. https://doi.org/10.4103/0019-5049.190620

Black, N. (2003). Secondary use of personal data for health and health services research: why identifiable data are essential. *Journal of health services research & policy*, *8*(1_suppl), 36-40.

Boslaugh, S. (2007). An introduction to secondary data analysis. *Secondary data sources for public health: A practical guide*, 2-10.

Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary use of EHR: data quality issues and informatics opportunities. *Summit on Translational Bioinformatics*, *2010*, 1.

Brakewood, B., & Poldrack, R. A. (2013). The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data. *Neuroimage*, *82*, 671-676.

Brown, I., Brown, L., & Korff, D. (2010). Using NHS patient data for research without consent. *Law, Innovation and Technology*, *2*(2), 219-258.

Burton, P. R., Banner, N., Elliot, M. J., Knoppers, B. M., & Banks, J. (2017). Policies and strategies to facilitate secondary use of research data in the health sciences.

Castle, J. E. (2003). Maximizing research opportunities: Secondary data analysis. *Journal of Neuroscience Nursing*, *35*(5), 287.

Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: opportunities and implementation. *Shanghai archives of psychiatry*, *26*(6), 371.

Coughlin, S. S. (2006). Ethical issues in epidemiologic research and public health practice. *Emerging themes in epidemiology*, *3*(1), 16.

Cox, E., Martin, B. C., Van Staa, T., Garbe, E., Siebert, U., & Johnson, M. L. (2009). Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects

using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value in Health*, *12*(8), 1053-1061.

Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013, May). Addressing big data issues in scientific data infrastructure. In *2013 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 48-55). IEEE.

Dolley, S. (2018). Big data's role in precision public health. *Frontiers in public health*, *6*, 68.

Edwards, L., & Harbina, E. (2013). Protecting post-mortem privacy: Reconsidering the privacy interests of the deceased in a digital world. *Cardozo Arts & Ent. LJ*, *32*, 83.

El Emam, K., Rodgers, S., & Malin, B. (2015). Anonymising and sharing individual patient data. *bmj*, *350*, h1139.

Endriyas, M., Alano, A., Mekonnen, E., Ayele, S., Kelaye, T., Shiferaw, M., & Hailu, S. (2019). Understanding performance data: health management information system data accuracy in Southern Nations Nationalities and People's Region, Ethiopia. *BMC health services research*, *19*(1), 175.

Floridi L. & Taddeo M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374* doi.org/10.1098/rsta.2016.0360

Floridi, L., Luetge, C., Pagallo, U., Schafer, B., Valcke, P., Vayena, E., & Vannieuwenhuyse, B. (2019). Key ethical challenges in the European medical information framework. *Minds and Machines*, *29*(3), 355-371.

Foley, B., Shuttleworth, I., & Martin, D. (2018). Administrative data quality: Investigating record-level address accuracy in the Northern Ireland Health Register. *Journal of Official Statistics*, *34*(1), 55-81.

Francis, B. (2020). General Data Protection Regulation (GDPR) and Data Protection Act 2018: What does this mean for clinicians?. *Archives of Disease in Childhood-Education and Practice*.

Frunză, A., & Sandu, A. (2017). Ethical acceptability of using generic consent for secondary use of data and biological samples in medical research. *Acta bioethica*, *23*(2).

Gatignon, H. (2019). Ethical behaviours versus behaviours that contravene deontological research principles in the publishing process. *Recherche et Applications en Marketing (English Edition)*, *34*(2), 63-74.

Godard, B., Schmidtke, J., Cassiman, J. J., & Aymé, S. (2003). Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective. *European Journal of Human Genetics*, *11*(2), S88-S122

Goroff, D., Polonetsky, J., & Tene, O. (2018). Privacy protective research: Facilitating ethically responsible access to administrative data. *The ANNALS of the American Academy of Political and Social Science*, *675*(1), 46-66.

Graves, A., McLaughlin, D., Leung, J., & Powers, J. (2019). Consent to data linkage in a large online epidemiological survey of 18–23 year old Australian women in 2012–13. *BMC Medical Research Methodology*, *19*(1), 235.

Gray, K., & Schein, C. (2012). Two minds vs. two philosophies: Mind perception defines morality and dissolves the debate between deontology and utilitarianism. *Review of Philosophy and Psychology*, *3*(3), 405-423.

Grinyer, A. (2009). The ethics of the secondary analysis and further use of qualitative data. *Social Research Update*, *56*(4), 1-4.

Hagger-Johnson, G., Harron, K., Gonzalez-Izquierdo, A., Cortina-Borja, M., Dattani, N., Muller-Pebody, B., & Goldstein, H. (2015). Identifying possible false matches in anonymized hospital administrative data without patient identifiers. *Health services research*, *50*(4), 1162-1178.

Haley, V. B., Van Antwerpen, C., Tserenpuntsag, B., Gase, K. A., Hazamy, P., Doughty, D., & Stricof, R. L. (2012). Use of administrative data in efficient auditing of hospital-acquired surgical site infections, New York State 2009–2010. *Infection Control & Hospital Epidemiology*, *33*(6), 565-571.

Hammond, W. E., Jaffe, C., & Kush, R. D. (2009). Healthcare standards development: The value of nurturing collaboration. *Journal of AHIMA*, *80*(7), 44-50.

Harpe, S. E. (2009). Using secondary data sources for pharmacoepidemiology and outcomes research. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, *29*(2), 138-153.

Harsh, R., Acharya, G., & Chaudhary, S. (2018). Epistemological View: Data Ethics, Privacy & Trust on Digital Platform. In *2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA)* (pp. 1-6). IEEE.

Hasan, S., & Padman, R. (2006). Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. In *AMIA annual symposium proceedings* (Vol. 2006, p. 324). American Medical Informatics Association.

Herbert, A, Wijlaars, L, Zylbersztejn, A, Cromwell, D, Hardelid, P. (2017). Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *International Journal of Epidemiology. Aug; 46*(4):1093-1093i. DOI: 10.1093/ije/dyx015.

Herschel, R., & Miori, V. M. (2017). Ethics & big data. *Technology in Society*, *49*, 31-36.

Heurix, J., & Neubauer, T. (2011). Privacy-preserving storage and access of medical data through pseudonymization and encryption. In *International Conference on Trust, Privacy and Security in Digital Business* (pp. 186-197). Springer, Berlin, Heidelberg.

Holzer, K., & Gall, W. (2011). Utilizing IHE-based electronic health record systems for secondary use. *Methods of information in medicine*, *50*(04), 319-325.

Hudson, K. L., & Collins, F. S. (2015). Bringing the common rule into the 21st century. *New England Journal of Medicine*, *373*(24), 2293-2296.

Huston, P., & Naylor, C. D. (1996). Health services research: reporting on studies using secondary data sources. *CMAJ: Canadian Medical Association Journal*, *155*(12), 1697.

Iacono, L. L. (2007). Multi-centric universal pseudonymisation for secondary use of the EHR. *Studies in health technology and informatics*, *126*, 239.

Iezzoni, L. I. (1997). Assessing quality using administrative data. *Annals of internal medicine*, *127*(8_Part_2), 666-674.

Information Commissioners Office (ICO). Exemptions. (2020). Retrieved 10/03/2020 from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/exemptions/

Iversen, A., Liddell, K., Fear, N., Hotopf, M., & Wessely, S. (2006). Consent, confidentiality, and the data protection act. *Bmj*, *332*(7534), 165-169.

Johnston, M. P. (2014). Secondary data analysis: A method of which the time has come. *Qualitative and quantitative methods in libraries*, *3*(3), 619-626.

Juddoo, S., & George, C. (2018). Discovering the most important data quality dimensions in health big data using latent semantic analysis. *IEEE International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD) Durban, South Africa, 6-7 August 2018*

Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., ... & Liaw, S. T. (2016). A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *Egems*, *4*(1)

Kalkman, S., Mostert, M., Gerlinger, C., van Delden, J. J., & van Thiel, G. J. (2019). Responsible data sharing in international health research: a systematic review of principles and norms. *BMC medical ethics*, *20*(1), 21.

Kaplan, B. (2014). How Should Health Data Be Used?: Privacy, Secondary Use, and Big Data Sales. *Cambridge Quarterly of Healthcare Ethics*, *25*(2), 312-329.

Lafoz, M. C., Ramírez-Soriano, A., & Richardson, S. (2018). Anonymisation: A new challenge for medical writers. *Medical Writing*, *27*, 31-36.

Lambert, T. W., Soskolne, C. L., Bergum, V., Howell, J., & Dossetor, J. B. (2003). Ethical perspectives for public and environmental health: fostering autonomy and the right to know. *Environmental Health Perspectives*, *111*(2), 133-137.

Langarizadeh, M., Orooji, A., & Sheikhtaheri, A. (2018). Effectiveness of Anonymization Methods in Preserving Patients' Privacy: A Systematic Literature Review. In *eHealth* (pp. 80-87).

Lawlor, D. A., & Stone, T. (2001). Public health and data protection: an inevitable collision or potential for a meeting of minds?.

Lowrance, W. (2003). Learning from experience: privacy and the secondary use of data in health research. *Journal of health services research & policy*, *8*(1_suppl), 2-7.

Martin-Sanchez, F. J., Aguiar-Pulido, V., Lopez-Campos, G. H., Peek, N., & Sacchi, L. (2017). Secondary use and analysis of big data collected for patient care. *Yearbook of medical informatics*, *26*(01), 28-37.

Mazeikiene, S., Stasiuniene, J., Vasiljevaite, D., Laima, S., Chmieliauskas, S., Fomin, D., & Jasulaitis, A. (2020). Deontological examination as a criterion for the assessment of personal healthcare professional quality: A Strobe compliant retrospective study. *Medicine*, *99*(3), e18770.

McLennan, S., Shaw, D., & Celi, L. A. (2018). The challenge of local consent requirements for global critical care databases. *Intensive care medicine*, *45*(2), 246-248.

Mészáros, J., & Ho, C. H. (2018). Big data and scientific research: the secondary use of personal data under the research exemption in the GDPR. *Hungarian Journal of Legal Studies*, *59*(4), 403-419.

Morrow, V., Boddy, J., & Lamb, R. (2014). The ethics of secondary data analysis: Learning from the experience of sharing qualitative data from young people and their families in an international study of childhood poverty.

Motulsky, A., Weir, D. L., Couture, I., Sicotte, C., Gagnon, M. P., Buckeridge, D. L., & Tamblyn, R. (2018). Usage and accuracy of medication data from nationwide health information exchange in Quebec, Canada. *Journal of the American Medical Informatics Association*, *25*(6), 722-729.

Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S. E., Bell, J., & Kaye, J. (2018). Are 'pseudonymised'data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review*, *34*(2), 222-233.

Nass, S. J., Levit, L. A., & Gostin, L. O. (2009). The value, importance, and oversight of health research. In *Beyond the HIPAA privacy rule: enhancing privacy, improving health through research*. National Academies Press (US).

National Academy of Sciences (US), N. (2009). Promoting the Stewardship of Research Data. Retrieved 11 March 2020, from: https://www.ncbi.nlm.nih.gov/books/NBK215270/?report=reader.

National Data Guardian for Health and Care's Review of Data Security, Consent and Opt-Outs. (2016). Retrieved 11 March 2020, from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF.

Neubauer, T., & Heurix, J. (2011). A methodology for the pseudonymization of medical data. *International journal of medical informatics*, *80*(3), 190-204.

Neumann, P. J., Sanders, G. D., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (Eds.). (2016). *Cost-effectiveness in health and medicine*. Oxford University Press.

Nikiforova, A. (2019). Analysis of Open Health Data Quality Using Data Object-Driven Approach to Data Quality Evaluation: Insights from a Latvian Context. In *IADIS International Conference e-Health* (pp. 119-126).

Peabody, J. W., Luck, J., Jain, S., Bertenthal, D., & Glassman, P. (2004). Assessing the accuracy of administrative data in health information systems. *Medical care*, 1066-1072.

Pezoulas, V. C., Kourou, K. D., Kalatzis, F., Exarchos, T. P., Venetsanopoulou, A., Zampeli, E., & Fotiadis, D. I. (2019). Medical data quality assessment: On the development of an automated framework for medical data curation. *Computers in biology and medicine*, *107*, 270-283.

Pieper, P. (2008). Ethical perspectives of children's assent for research participation: deontology and utilitarianism. *Pediatric nursing*, *34*(4), 319-324.

Ploug, T., & Holm, S. (2015). Meta consent: a flexible and autonomous way of obtaining informed consent for secondary research. *Bmj*, *350*, h2146.

Pommerening, K., & Reng, M. (2004). Secondary use of the EHR via pseudonymisation. *Studies in health technology and informatics*, 441-446.

Pormeister, K. (2017). Genetic data and the research exemption: is the GDPR going too far?. *International Data Privacy Law*.

Porsdam Mann, S., Savulescu, J., & Sahakian, B. J. (2016). Facilitating the ethical use of health data for the benefit of society: Electronic health records, consent and the duty of easy rescue. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 20160130.

Prada-Ramallal, G., Takkouche, B., & Figueiras, A. (2019). Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review. *BMC medical research methodology*, *19*(1), 53.

Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature medicine*, *25*(1), 37-43.

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J. C., & Ghali, W. A. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical care*, 1130-1139.

Resnik, D. B. (2015). Paternalism and utilitarianism in research with human participants. *Health Care Analysis*, *23*(1), 19-31.

Ross, M. W., Iguchi, M. Y., & Panicker, S. (2018). Ethical aspects of data sharing and research participant protections. *American Psychologist*, *73*(2), 138.

Rumbold B, Lewis G and Bardsley M (2011) Access to person-level data in health care: Understanding information governance. Research summary. Nuffield Trust. Retrieved 11 March 2020, from https://www.nuffieldtrust.org.uk/research/access-to-person-level-data-in-health-care-understanding-information-governance?gclid=Cj0KCQjw0pfzBRCOARIsANi0g0sAyThacXsmmIihOKclVFN3vnNmeI0PaC9H2WMQjymqiGWYkD6n3CUaAkuoEALw_wcB

Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., & Detmer, D. E. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association*, *14*(1), 1-9.

Schlegel, D. R., & Ficheur, G. (2017). Secondary use of patient data: review of the literature published in 2016. *Yearbook of medical informatics*, *26*(01), 68-71.

Singleton, P., & Wadsworth, M. (2006). Consent for the use of personal medical data in research. *Bmj*, *333*(7561), 255-258.

Somolinos, R., Muñoz, A., Hernando, M. E., Pascual, M., Cáceres, J., Sánchez-de-Madariaga, R., & Salvador, C. H. (2014). Service for the pseudonymization of electronic healthcare records based on ISO/EN 13606 for the secondary use of information. *IEEE journal of biomedical and health informatics*, *19*(6), 1937-1944.

Stalla-Bourdillon, Sophie, & Knight, Alison. (2016). Anonymous data v. personal data - a false debate: An EU perspective on anonymization, pseudonymization and personal data. Wisconsin International Law Journal, 34(2), 322.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, *40*(5), 103-110.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557-570.

Takahashi, A., Kumamaru, H., Tomotaki, A., Matsumura, G., Fukuchi, E., Hirata, Y., & Miyata, H. (2018). Verification of data accuracy in Japan congenital cardiovascular surgery database including its postprocedural complication reports. *World Journal for Pediatric and Congenital Heart Surgery*, *9*(2), 150-156.

Taleb, I., El Kassabi, H. T., Serhani, M. A., Dssouli, R., & Bouhaddioui, C. (2016). Big data quality: A quality dimensions evaluation. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)* (pp. 759-765). IEEE.

Taylor, P. (2013). Caldicott 2 and patient data. *BMJ 2013; 346: f2260*

Tenenbaum, J. D., Avillach, P., Benham-Hutchins, M., Breitenstein, M. K., Crowgey, E. L., Hoffman, M. A., & Ray, B. (2017). An informatics research agenda to support precision medicine: seven key areas. *Journal of the American Medical Informatics Association*, *23*(4), 791-795.

Thomson, D., Bzdel, L., Golden-Biddle, K., Reay, T., & Estabrooks, C. A. (2005). Central questions of anonymization: A case study of secondary use of qualitative data. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 6, No. 1).

Tripathy, J. P. (2013). Secondary data analysis: Ethical issues and challenges. *Iranian journal of public health*, *42*(12), 1478.

Tu, K., Campbell, N. R., Chen, Z. L., Cauch-Dudek, K. J., & McAlister, F. A. (2007). Accuracy of administrative databases in identifying patients with hypertension. *Open medicine*, *1*(1), e18.

Van Mourik, M. S., van Duijn, P. J., Moons, K. G., Bonten, M. J., & Lee, G. M. (2015). Accuracy of administrative data for surveillance of healthcare-associated infections: a systematic review. *BMJ open*, *5*(8), e008424.

Vartanian, T. P. (2010). *Secondary data analysis*. Oxford University Press.

Verheul, E. R., Jacobs, B., Meijer, C., Hildebrandt, M., & de Ruiter, J. (2016). Polymorphic Encryption and Pseudonymisation for Personalised Healthcare. *IACR Cryptology ePrint Archive*, *2016*, 411.

Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*, *20*(1), 144-151.

Weiskopf, N. G., Hripcsak, G., Swaminathan, S., & Weng, C. (2013). Defining and measuring completeness of electronic health records for secondary use. *Journal of biomedical informatics*, *46*(5), 830-836.

Wilkerson, M. L., Henricks, W. H., Castellani, W. J., Whitsitt, M. S., & Sinard, J. H. (2015). Management of laboratory data and information exchange in the electronic health record. *Archives of Pathology and Laboratory Medicine*, *139*(3), 319-327.

Windle, P. E. (2010). Secondary data analysis: is it useful and valid?. *Journal of PeriAnesthesia Nursing*, *25*(5), 322-324.

Wise, J. (2019). Price hike makes access to patient data unaffordable, say researchers. *BMJ*, 2019; 366: l5305

Xafis, V., Schaefer, G. O., Labude, M. K., Brassington, I., Ballantyne, A., Lim, H. Y., & Laurie, G. T. (2019). An ethics framework for big data in health and research. *Asian Bioethics Review*, *11*(3), 227-254.

Xiao, Y., Bochner, A. F., Makunike, B., Holec, M., Xaba, S., Tshimanga, M., & Feldacker, C. (2017). Challenges in data quality: the influence of data quality assessments on data availability and completeness in a voluntary medical male circumcision programme in Zimbabwe. BMJ open, 7(1), e013562.

--0--

How to cite this article

Al Omari, M. (2021). Using Secondary Health Data in Research. *Malaysian Journal of Medical and Biological Research*, *8*(1), 31-40. https://doi.org/10.18034/mjmbr.v8i1.544