

Ethics Considerations for Conducting Secondary Data Analyses

Tony Tan

Centre for Educational Measurement, University of Oslo

UV9010 Research Ethics

Associate Prof Anett Kaale

15 November 2021

Ethics Considerations for Conducting Secondary Data Analyses

Fairness, equality and equity are central concerns in educational sciences and practices. Unlike their counterpart in the USA, universities in Norway make admission decisions exclusively based on academic achievement. This practice brings the fairness of the grade point average (GPA) system into sharp focus. In particular, differences in GPA subject difficulties inevitably create disincentives and distortionary redistribution of social resources, leading to a weakened outcome in educational fairness. My PhD project aims to quantify such inter-subject difficulties using Norwegian registry data and to propose a corrective statistical procedure to realign GPA subjects to a common difficulty scale.

Conducting quantitative research using existing data sets raises unique ethical considerations. Similar studies by He and Stockford (2015) and He et al. (2018) in the UK as well as Korobko et al. (2008) in the Netherlands have demonstrated the attractiveness of secondary analyses using archival data administered by the education authorities. The ethical dilemma of repurposing and combining data sets, however, has been understudied particularly in the quantitative branch of educational sciences. This paper wishes to address such gaps by firstly overviewing the ethical guidelines and procedures relevant to secondary analyses (SA) particularly the dilemma for informed consent; it then links the ethics framework with a cost-benefit analysis of SA as a research methodology.

Ethics Principles, Informed Consent, and Reidentification

It is chiefly important to distinguish ethics considerations from legal requirements. Both ethics and laws serve to safeguard individual dignity and advance social well-being but they operate at different standards of proof and with various enforcement power. While ethics are a set of moral values an individual establishes for themselves and their own personal behaviour, laws are structured rules aimed to govern all of society. In the research profession, legal requirements prescribe the minimum standard of conduct whereas ethical guidelines strive for maximum goodness for all parties involved with heavy emphases on the vulnerable such as study participants.

Resultantly, this paper would not focus on the general data protection regulation (GDPR) legal framework in favour of the overarching considerations over respect for persons through informed consent.

In the Norwegian context, researchers of social sciences and humanities may seek ethics guidelines and advisory from *Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora* (NESH, 2018), whose positions and philosophies are largely comparable to their counterparts in the USA, from where many of this paper's citations are drawn. The ethical underpinnings of human research participant protection in Norway are firmly grounded in the Western philosophical traditions that uphold *individual rights* as paramount. Three ethical aspects are particularly salient for educational research: (a) respect for persons, (b) beneficence, and (c) justice. The first principle requires recognising individuals as autonomous agents with unalienable rights for self-determination. Additional protection must be afforded to those possessing less capacity of self-determination such as minors and consumers of special education programs (Ross et al., 2018). Researchers shall demonstrate their respect for persons in all study projects by presenting prospective participants with opportunities to give *voluntary* and *informed* consent. Such invitation must be sufficiently full and accessible for the participants to enable their decision about whether to take part. Prospective participants then must be allowed to decide freely either to take part, to decline, or to withdraw without attracting adverse consequences from them (Crow et al., 2006). The second principle beneficence addresses the researchers' obligation to protect the well-being of human participants by minimising risks of harm and maximising any potential benefits to the participants. Should the explicit benefits to the individual participant be minimal, the studies can be ethically acceptable if substantial societal benefits exist and given that every effort is made to protect the participants from risks of harm (Ross et al., 2018). Lastly, the justice principle addresses equitable distribution of benefits and burdens of research on humans. It requires, as a methodological example, that a fair selection of research participants to be based on the missions and hypothesised outcomes of the research, and not based on easy availability or

manipulability of the recruitment pool (Ross et al., 2018).

Secondary analyses drawing Norwegian students' GPA records are vulnerable to all three principles. Accessing, extracting, and aggregating archival data has profound impact on the practical interpretation of the principle of respect for persons, forcing a reassessment of the fundamental concept of voluntary and informed consent as the bedrock of ethically responsible research. Information deposited in Norway's national registry will likely be accessed by researchers other than the team that collected the data, and for purposes that might differ significantly from the foci of the original administrative designs. Consequently, young Norwegians, many below the age of 18, will not have control over who might access their academic performance data in the future and for what purposes as these were unknown factors at the initial consent stage, neither can they feasibly prevent their records from being accessed and studied after the de-identification process. That is, SA research is not based on informed consent in the traditional sense (Ross et al., 2018) but on open-ended and/or broad consent to an unspecified range of future research projects (Grady et al., 2015).

Furthermore, reidentification imposes material risks of harm to registry data participants. Although all information in the Norwegian national register is de-identified, such process is becoming increasingly insufficient for protecting individual identities due to cheap and rapid data aggregation and triangulation capabilities equipped by recent technological advancement. A 2015 study by de Montjoye et al., for example, has successfully reidentified 90 per cent of individual consumers using only four spatial-temporal points from their credit card metadata, leading to a caution from the authors that even apparently anonymous data, such as economic, behavioural, and spatial-temporal information, can be rendered identifiable once aggregated.

Re-identification risks are even higher in Norwegian registry data sets due to preservations of administration identifiers such as school and region codes.

Reidentification directly interrogates the current ethical and legal standards that exempt secondary uses of deidentified data from reporting obligations, positioning SA on an equal footing with the original data collectors under the beneficence principle.

Such understanding also imposes great burden of proof on the SA researchers to demonstrate significant social gains to future GPA candidates while committing to non-disclosure obligations at every stage of the analyses to protect the identity of earlier year high school graduates.

A Cost-benefit Analysis of Secondary Data Research

None of the proceeding paragraphs, however, shall be interpreted as an attack against research using secondary data sources. Good science must enable falsifiability by opening its data for scrutiny. Secondary analyses reinforce open science enquiry by replicating, reanalysing and evaluating prior results (Hedrick, 1985). Advancement in understanding of the ætiology of diseases, for example, often occurred through secondary analyses of medical records that were not collected with the intention of making such a causal inference (Dale et al., 1988). In addition, reuse of data sets promotes positive externality to society through its cost-effectiveness and convenience. Since Norwegian residents collectively finance the compilation of registry data through taxation, such information shall remain public property in principle (Davey Smith, 1994). Secondary analyses using registry data therefore serve to maximise social benefits through additional learning, knowledge and insight for the betterment of social well-being. Thirdly, secondary data analyses also serve to reduce harm to society. This is achieved by reducing the impact on the larger population by involving a smaller number of research participants and subjecting them to a smaller number of tests while a significant multiple of the original test subjects benefit from the resulting scientific inferences at zero additional costs (Law, 2005). Finally, since an enhanced methodology is one primary mechanism for reducing harm, the reemployment of prior data sets and studies assists subsequent projects with more targeted research questions and more purposeful analysis strategies (Davey Smith, 1994). SA hence carry enormous potential for advancing scientific understanding at minimal marginal costs to the researchers.

The incremental costs to the participants, on the other hand, may not be comparable to that born by the researchers due to power imbalance. Obtaining informed consent for secondary research is never truly voluntary due to the

foot-in-the-door effect (Freedman & Fraser, 1966) such that participants are vulnerable to giving away consent to subsequent projects after agreeing to the original study. Furthermore, Law (2005) observed that although it is the *participants'*, not the researchers', interpretation of non-disclosure of personal data that all ethical considerations prioritise, the former is still under the influence of the latter while making such interpretation. Compounded with logistical impracticality, few scholars or regulatory bodies advocate for consent renewals as a solution for the ethical dilemma facing SA research. Educational scientists, fortunately, do have the privilege of learning from related fields such as medicine where secondary analyses of biological specimens also trigger ethical dilemmas. As an alternative to renewed consent, Grady et al.'s (2015) proposal of obtaining "broad consent" from biospecimen donors was critically received by behavioural researchers due to concerns that blanket consent requirement would skew participation such that only data of those who had agreed to broad/blanket consent would be available for SA research. This unsatisfactory status of affair has highlighted the nature of research ethics: unlike in the judicial world, no clear-cut decisions can be reached, let alone inter-transplanted, amongst ethics debates. Researchers must remain alert to dilemmas, exercise good judgement at all times both as professionals and as responsible members of humanity, in bringing about well-balanced solutions to the ever-changing challenges presented to them.

Conclusion

An academic culture that centres around ethics considerations serves both the research and the researchers. Visible conversations about ethical conducts condition both graduate students and experienced mentors into reflective practice—a mental process that forces researchers to clarify what their project is intended to achieve, how those objectives can be best pursued, at what costs and to whom, under whose powers and privileges, and most importantly, does it promote the wellness of the society as a whole (Crow et al., 2006). This paper focuses on one particular field of scientific methodology of secondary data analyses and critically reflected the ethics considerations associated with reusing Norwegian GPA records for the purpose of

extracting educational fairness measures. Although secondary analyses of existing data sets may provide many exciting opportunities for the development of new knowledge, advancement in technology has also challenged the sufficiency of anonymisation as protection for participants' identities due to ease of data aggregation, triangulation and reidentification. Researchers and regulatory bodies are still in the process of balancing the sizeable marginal social gains against the potential risks of harm to the GPA record "donors". Both renewed consent and broad consent approaches have been evaluated for their practicality, ethical implications, as well as distortionary sampling effects for subsequent inferences. This paper reaches a position that it may not be the "what to do" but "why do so" that shall guide all researchers in making their scientific decisions. May my projects advance the well-being of future graduates through a fairer GPA design.

References

- Crow, G., Wiles, R., Heath, S., & Charles, V. (2006). Research ethics and data quality: The implications of informed consent. *International Journal of Social Research Methodology*, 9(2), 83–95. <https://doi.org/10.1080/13645570600595231>
- Dale, A., Arber, S., & Procter, M. (1988). *Doing secondary analysis: A practical guide*. Routledge.
- Davey Smith, G. (1994). Increasing the accessibility of data. *British Medical Journal*, 308, 1519–1520. <https://doi.org/10.1136/bmj.308.6943.1519>
- Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 4(2), 195–202. <https://doi.org/10.1037/h0023552>
- Grady, C., Eckstein, L., Berkman, B., Brock, D., Cook-Deegan, R., Fullerton, S. M., Greely, H., Hansson, M. G., Hull, S., Kim, S., Lo, B., Pentz, R., Rodriguez, L., Weil, C., Wilfond, B. S., & Wendler, D. (2015). Broad consent for research with biological samples: Workshop conclusions. *The American Journal of Bioethics*, 15(9), 34–42. <https://doi.org/10.1080/15265161.2015.1062162>
- He, Q., & Stockford, I. (2015). *Inter-subject comparability of exam standards in GCSE and A Level* (ISC Working Paper 3). Office of Qualifications and Examinations Regulation. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/606044/3-inter-subject-comparability-of-exam-standards-in-gcse-and-a-level.pdf
- He, Q., Stockford, I., & Meadows, M. (2018). Inter-subject comparability of examination standards in GCSE and GCE in England. *Oxford Review of Education*, 44(4), 494–513. <https://doi.org/10.1080/03054985.2018.1430562>
- Hedrick, T. E. (1985). Justifications for and obstacles to data sharing. In S. E. Fienberg, M. E. Martin, & M. L. Straf (Eds.), *Sharing research data* (pp. 123–147). National Academies Press. <https://doi.org/10.17226/2033>
- Korobko, O. B., Glas, C. A. W., Bosker, R. J., & Luyten, J. W. (2008). Comparing the difficulty of examination subjects with item response theory. *Journal of*

Educational Measurement, 45(2), 139–157.

<https://doi.org/10.1111/j.1745-3984.2007.00057.x>

Law, M. (2005). Reduce, reuse, recycle: Issues in the secondary use of research data.

IASSIST Quarterly, 29(1), 1–6. <https://doi.org/10.29173/iq599>

de Montjoye, Y.-A., Radaelli, L., Singh, V. K., & Pentland, A. S. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *347*(6221),

536–539. <https://doi.org/10.1126/science.1256297>

NESH. (2018). *Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi*. De nasjonale forskningsetiske komiteene.

<https://www.forskningsetikk.no/retningslinjer/hum-sam/forskningsetiske-retningslinjer-for-samfunnsvitenskap-humaniora-juss-og-teologi/>

Ross, M. W., Iguchi, M. Y., & Panicker, S. (2018). Ethical aspects of data sharing and research participant protections. *American Psychologist*, 73(2), 138–145.

<https://doi.org/10.1037/amp0000240>