Extending the Uppsala Birth Cohort Multigenerational
Study Database With a New Collection From Archival
Sources: Collection and Error Correction Strategies
When Problems Arise

Contributors: Anthony M. Garcy

Pub. Date: 2018

Access Date: January 8, 2018

Academic Level: Postgraduate

Publishing Company: SAGE Publications Ltd

City: London

Online ISBN: 9781526440013

DOI: <http://dx.doi.org/10.4135/9781526440013>

©2018 SAGE Publications Ltd. All Rights Reserved.

This PDF has been generated from SAGE Research Methods Cases.

Abstract

This case study is an account of the 2014-2016 effort to expand a Swedish research database called the Uppsala Birth Cohort Multigenerational Study. The research project collected and photographed available data on school quality from local, regional, and national Swedish archives. The discovery of a widespread data quality issue in the existing database ultimately prevented the completion of the data collection and the execution of the planned research. A narrative is given about the challenges of conducting a complex, multistage archival data collection. Some of the problems that were encountered are mentioned. Practical methods and strategies that were used to collect the relevant data from the archival material are discussed. The methods used in the conversion and entry of some of this material into an electronic, numerical database format are also reviewed.

Learning Outcomes

By the end of this case, students should be able to

- Understand why it is important to have a systematic plan informed by a pilot survey before the full data collection in the archives begins
- Analyze this case study and compare the archival data collection methods that were used at different points in time
- Suggest what strategies could have been used to address the problems that were created by the methods employed by the original archival collections
- Evaluate how a researcher might increase the reliability of a large, complex, archival data collection

Project Overview and Context

At the beginning of 2014, my research team and I began with the ambitious task of expanding a well-used Swedish research database called the Uppsala Birth Cohort Multigenerational Study (UBCoS). A number of studies have been published from these data (e.g., Gisselmann, Koupil, & De Stavola, 2011; Modin, Erickson, & Vågerö, 2012; Van den Berg & Modin, 2013). Ultimately, the discovery of a widespread data quality issue prevented the completion of this task. Despite this problem, we nearly succeeded.

This case study provides a narrative about the challenges of conducting a complex, multistage archival data collection and some of the difficulties that were encountered. I describe the practical methods and strategies that were used to collect the relevant data from the archival material. The methods used in the conversion and entry of some of this material into an

electronic, numerical database format are also mentioned. This case study will not review the basics of research that is often conducted in the archival setting. A very informative SAGE case study is currently available that addresses this topic (see Wangard, 2016).

The UBCoS database includes five generations of related individuals and their partners ($n = 140,000$). It is possible to study how the early life characteristics and social circumstances of these individuals influenced the later health and social outcomes of the original cohort members (Generation 1—G1) and their descendants. This nationally representative G1 cohort consists of more than 14,100 Swedish individuals. They were born in the Uppsala University Hospital during the period 1915-1929. Longitudinal social and health information spanning much of the life course of these cohort members has been previously collected from archival and contemporary registry sources. Birth characteristics, information about early school achievement (third-grade school marks in 12 subjects), later educational attainment, and information about parents and family were collected from church parish records, school archives, and archived census records. This information has also been linked to later contemporary registers kept by Swedish state statistical agencies. In 2004, the original G1 birth cohort was linked to the social and health data of all existing descendants.

In late 2013, I was awarded funding from the Swedish Research Council to investigate a timely but previously unaddressed research question: Did higher school quality early in the life course modify the well-known negative effects of disadvantageous birth characteristics (e.g., low birth weight) on a variety of later social and health outcomes? The UBCoS database contained much of the information needed to address this question. However, information on the school quality of the cohort members had never been collected. Based on the findings of two earlier pilot surveys, my team and I had determined that this type of information was available in the Swedish archives. School quality information, contained on standardized paper forms, included data on the cohort member's class sizes, resources spent in the classroom, teacher certification, teacher experience, and pay levels. Other administrative financial information that could be used to construct a variety of school quality measures was also available. Importantly, the detail of this archival information on school quality rivals similar contemporary register information available from Swedish state statistical agencies. The pilot surveys and our previous knowledge about the time it had taken to collect the original archival school achievement data for individuals (school marks) allowed us to estimate the amount of time our new collection would take (2-2.5 years). This was critical given the limited project funding period.

Research Practicalities

It was evident to me that the scope of the task of collecting and adding this new school quality

information to the existing UBCoS database would be very large and complicated. A series of staggered collections of some school information including marks and school names for this G1 cohort had been previously done over several years. Data were collected from various local, regional, and national archives located all over Sweden. The collection first began in Uppsala County and then continued elsewhere. Although the majority of the G1 cohort had attended school in Uppsala County in 1924-1940, almost a quarter of the cohort with subject marks attended school elsewhere in Sweden. Thus, the school quality information for these cohort members was located in archives in other areas outside of Uppsala County. This presented a logistical and a financial challenge. Hence, we developed a systematic, staggered collection plan to ensure the likelihood that we could obtain as much of the data in the shortest period of time for the lowest cost.

In the first year, we began the collection in the Uppsala County archives where the majority of the information was located. In the second year of the collection, we proceeded to Stockholm County where the next largest concentration of archival material was stored. In the final stage, we focused our collection efforts on other geographical areas where there were large clusters (10 or more) of cohort members. For areas where we had only a small number of cohort members (nine or less) and a physical visit was impractical due to great distance and high travel cost from Stockholm, we directly requested the information from the archivist via a detailed email. We began contacting these archives at the beginning of the second year of the collection. Multiple follow-ups with these distant archives was typical.

An additional complexity concerned the methods used in the original collections. A major shortcoming was that no permanent record of the original school archival material that the database was created from existed. In the original collection that began in Uppsala County, school information that was needed was first written by hand on pre-printed paper forms. Other potentially useful information (e.g., the student's teacher name) that was available in the original archival material had not been recorded. This posed a problem for us because the new school quality information that we were going to collect had to be linked to the existing school marks data according to the cohort member's teacher name. To obtain the teacher name, it was evident that we would also have to collect the original archived grade registers for nearly 10,500 G1 cohort members in addition to all of the information I wished to add.

A second substantial shortcoming of the original Uppsala County collection was high key entry error. In a number of Uppsala County districts, we found key error entry rates in excess of 271 per 10,000 entries (Garcy, 2016). Expected single entry error rates have been previously reported at 54 to 72 per 10,000 entries (Büchele, Och, Bolte, & Weiland, 2005; Williams, Garrett, & Petroni, 2004). The key entry error rate in some school districts was at least four

times these published rates! Given the possibility for data entry errors to occur when transposing the information from the archival material to the paper forms and then from the paper forms to the electronic database (Wahi, Parks, Skeate, & Goldin, 2008), the key entry error rate had likely been multiplied.

A third more problematic shortcoming of the later stage of the original collection included a set of widespread systematic errors in the school marks data that minimally affected 33% of the collected cases in school districts outside of Uppsala County (Garcy, 2016). These errors were introduced into the database as a result of a change in the data entry method. Instead of continuing with the handwritten paper data collection forms and later separate data entry, the information in the archival material was directly entered into an electronic database on site. This shift to a direct entry method should theoretically have decreased key entry errors. However, individual data files were typically created in each archive and then later merged into a single file. But database management software was not used. As a result, these files were not merged properly. Rather, it was likely that some of the data were cut and pasted from each individual file into a central master file. Unfortunately, the individual data fields in each of the separate files were not typically in the same order. Hence, subject marks were frequently switched with each other in different districts (e.g., math with the Swedish language) when the individual district files were merged together. Again, there was no permanent record of the original archival material. So, there was no way to check the reliability of the master database after it had been assembled.

Research Design

When we began the new collection, we knew nothing of these widespread problems with the quality of the school marks data. It was well over a year into the new collection when they were first discovered. We adjusted our collection strategy along the way to allow for the possibility to fix them because these data were ultimately needed to address the central research question mentioned earlier.

At the outset of the project and in anticipation of the complexity of the proposed collection, we were acutely aware of the need to have a set of formalized collection and data entry methods to ensure high reliability. Hence, we developed a detailed plan to create a permanent record of the original documents by digitally photographing them. It was reasoned that we or other researchers might wish to extract other information from this material in the future. We also created a system for organizing all of the digital photographs as we anticipated generating a very large number of images. A third part of the plan involved double data entry of the needed information from the digital photographs into an electronic database. This was to be done at

regular intervals in a separate process from the collection in the archives. The aim was to reduce the key entry error rate and produce the highest reliability of the new school quality data.

Method in Action

Two project assistants were hired in the first year of the collection. At the beginning, I could not be personally involved in the daily archival visits. One of the assistants had significant, previous first-hand experience with the later part of the original collection. Her understanding of the organizational structure of the archives, her familiarity with the archival personnel, and her knowledge of the material we intended to collect ensured that we could adhere to the collection schedule. This expertise and experience were critical and invaluable. In fact, if I had not been able to employ such a person at the outset, I would not have moved forward with the project.

The assistants began photographing the original grade registers and the new school quality material in early 2014. Most of the material was found in two central locations in Uppsala County. I first considered using an expensive, large SLR (single lens reflex) digital camera with a tripod to ensure the highest quality and most consistent images possible. This type of system is often used by archives that are in the process of digitalizing their collections for online access. As a result of the earlier pilot surveys to the archives, I had noted how cumbersome these systems were in practicality. Replacement was also a consideration in case the equipment was broken, lost, or stolen.

As our aim was to be able to reference the archival material at a later point, I opted instead to use the highly rated, relatively inexpensive, compact point and shoot Canon PowerShot SX280 HS digital camera. I had experimented with the camera and determined that it would perform very well under a variety of lighting conditions in the archives. The quality of the test images was very good even when magnified on a computer screen. Finally, it could be used for multiple hours without significant user fatigue as it was small and was lightweight.

I gave the assistants standardized instructions on their use before the collection began. The cameras were always set to a standard 2816 × 1584 pixel resolution and were used in either auto or portrait mode to ensure that the quality of the images would be very high. The cameras had an autofocus function with audible and visual indicators to increase the likelihood that images would be in focus. The cameras had some manual settings (shutter and aperture) that could have been adjusted in a low light setting. We did not use the flash as it reduced operating time and was generally not needed. Lighting in the archives was almost always adequate. In addition, many archives will only allow photographing of the material with the

natural lighting to avoid damage to the documents.

When we began the new collection, we did not have a list of the G1 cohort members' names. But we did have a complete list of the districts and schools they had attended and the year in which the cohort members had attended the school. Hence, all third-grade registers in a district were collected for the specific school and specific attendance year that was needed. In fact, this was a very efficient collection strategy because we usually found that groups of cohort members were enrolled in the same third-grade classes in each school. To relocate every individual student cohort member separately would have been very time-consuming, especially where the majority of a class comprised cohort members. This was the case in many Uppsala County districts.

After the grade registers were collected, the cohort member's teacher name was identified. The assistants then collected the corresponding school quality information which was usually found on standardized archived paper forms. A rudimentary organizational scheme was first followed where all digital photographs that had been taken in a day were first uploaded to folders on a laptop according to the name of the school district and the date taken. A daily backup copy was also made to an external, encrypted hard drive. This proved to be very important later when one of the laptop computers failed. This organizational scheme proved to be inadequate later in the early summer of 2015 when the first wave of intensive data entry began after the discovery of the data quality issue.

Shortly after the archival material had been photographed, the relevant data were entered into an electronic database format by each project assistant individually. This was done at a remote location away from the archive to provide an optimal setting for reliable data entry. The results of these individual entry efforts were then compared and discrepancies were noted and resolved as the original material could easily be re-referenced because it had been photographed. This obviously could not have been done in the original collection. This double data entry method, while more time-consuming, has been shown to reduce key entry error rates that are substantially lower than single key entry error rates (Gibson, Harvey, Everett, & Parmar, 1994; Reynolds-Haertle & McBride, 1992).

Toward the end of the first year of the new collection, the second phase began. Almost all of the Uppsala County archival material had been photographed by the two assistants. As planned, they moved the collection to archives mostly located in Stockholm County. It was at this time that one of the assistants decided to leave for the United Kingdom. A replacement was hired and trained by both of the original assistants. At the same time, as originally planned, I also became involved in the daily collection as the demands of the more geographically

dispersed search had increased. I could also closely monitor the new assistant and ensure that collection protocol was followed and address any problems that arose. This involvement proved to be crucial in the discovery of the existing database quality problems.

From the start of the second phase of the new collection, we had obtained a complete list of all the cohort members' names, each school district and school name, the year each student attended the third grade, and their school marks in each of the 12 subject areas. This list was critical. It allowed us to conduct the collection in a prescribed way as the number of students we were searching for was much smaller when compared with typical schools located in the Uppsala County districts.

About a month into our collection in the Stockholm County archives, my recently hired project assistant and I began to notice discrepancies between the school marks that had been entered into the master database and the school marks that were found in the relocated archival material. At first, this did not raise any concern as I assumed these errors were attributable to very careless data entry. But as we continued the collection in different archives, we found several districts where most of the student school marks did not match those that had been entered beforehand. By the early spring of the 2015, we had collected data in 15 districts located in different archives where match rates were less than 20%. At this point, it was evident to me that we had discovered a widespread systematic error in the database. Naturally, I became very alarmed.

I notified the Primary Investigation (PI) of the UBCoS study and hoped to receive some reasonable explanation for what we had found. The PI's first response was surprise and some disbelief. But she was quickly persuaded that a problem existed after she was shown some of the original archival material compared with the subject marks in the master database. However, she had no explanation for what we had discovered.

One of the benefits of having photographed the material from each archive was that we were able to compare it. In doing so, we had noticed from these pictures that the printed format of the original grade registers had often changed over time in different districts. As a result, the ordering of the subject marks was often different from year to year. Different districts had used different printers and they even changed printers over time. We also noted other discrepancies in some of the school data contained in UBCoS when it was compared with the photographed material. These included inaccuracies in the recorded number of absences for students and numerous mistakes in their school names. These errors, considered together with the widespread errors in the subject marks, led to the development of an overall sense that the original collection had not reliably recorded much of the information from this archival material.

In a very short time span, the new collection and expansion of the UBCoS database had an additional, unexpected task. I was now in the position where I would have to fix the underlying school data before I could use it in any analyses. This was a major setback to which I only saw one solution.

At the start of the summer of 2015, we began an intensive 3-month attempt to identify the sources of the errors and potentially correct them. We developed a plan that first involved re-entry of some of the school marks data from the recently collected archival data photos. The main focus would be on the districts with the highest level of error. We also included a number of other very large districts within Uppsala County to check if problems existed there as well.

To implement this plan required substantial effort to organize the material we had photographed so that we could first find the relevant cohort members. Although we already had a rudimentary scheme in place to organize all of the photos that had been taken, it was not robust enough for the “fix attempt.” We developed a set of rules for creating a file structure and file folder naming scheme. Then, we sorted and organized the material according to this file structure and naming convention. Photos were organized by the district, then by the schools within the district, and then by the school year the cohort member had attended the third grade. This made the later process of searching the photographs and conducting the comparison with the master database as efficient as possible.

After the photos had been sorted and organized, we used the lists to find the cohort members within them. They were marked and saved with digital visual indicators using photo editing software. For example, we drew small red arrows next to the cohort member’s names and drew red boxes around the relevant school marks. This was done so that the cohort member’s information could be easily identified during the data entry process. Another benefit was that they could be found quickly in the photos at a later time. Because this marking process in the photographs was time-consuming, I decided from that point forward that we would change our method in the archives. We began to use place markers, positioned on the material prior to photographing it to indicate where the cohort member had been found.

By the summer’s end, our efforts had produced a file that made corrections to the master file in the four subject areas of math, reading, Swedish, and writing. We gave this file to the database administrator of UBCoS to implement these corrections. In a later summary meeting, I was given a brief explanation about the source of the original error. The database administrator asserted that one of the original files used to construct the master UBCoS database was found to be inexplicably missing the column variable names. Evidently, at a later point, a version of this same file had been assigned the wrong column variable names for nine of the 12 school

subjects. The database administrator had arrived at this conclusion by comparing a small number of cases in the pictures we had taken to the same cases in the master file. The database administrator stated that the correct variable name ordering had been applied to the corrupted file and a new “fixed” master file had been created.

I did not know this at the time, but not all of our corrections had been integrated into the new master file after the first fix attempt. Based on a statistical analysis conducted a year later, I determined that many errors we had found that were not attributable to the missing variable column names had not been dealt with at all. It also appeared that another unidentified file was potentially corrupted; there were other sources of error, or both.

Within a month of the end of the summer of 2015, we were collecting data in another set of community archives. Unfortunately, we saw the same problem of non-matching school marks for the majority of the cohort members in several districts. In other districts, we found the school marks data to be mostly correct. Although the problem was not as widespread as it had first been, the new districts we found with high levels of non-matching school marks proved to be systematic.

Based on the fact that the original characterization of the data corruption was not completely accurate, it had become clear to me that every district in the database would have to be checked and compared with the pictures that were collected. This was the only way to assure that the error level would be minimized to an acceptable level. In a later UBCoS steering group meeting, members suggested that they were comfortable with an error rate of 2% to 3% in the affected file. An estimate from my own analysis placed the minimal level of remaining error, after the first fix attempt, to be at least 7% (Garcy, 2016).

In early October 2015, I raised the issue with the UBCoS study PI again and related the fact that we had found a number of additional districts with widespread errors that had not been resolved despite our recent 3-month effort. I also relayed the concern that given this finding we could expect that other problematic districts would be found as the collection continued. The UBCoS PI suggested that we could try to “fix” the database again after we had collected additional data.

We discussed another attempt in the early winter of 2015, but the demands of the collection intensified and an immediate effort was not possible. Moreover, I reasoned that it would be a more efficient use of resources to finish the collection of the remaining material first so that we would not miss any heavily corrupted districts. An even greater problem was that the project had no budgeted resources for a new effort or for the earlier summer fix attempt we had

already implemented. I mentioned this and said that I could not divert additional resources again as I was concerned that the diversion of time and resources for the first fix attempt had already hindered the project. To do so again would almost certainly jeopardize the other parts. I asked whether resources from one of her other projects could be used to fund the new fix attempt. Unfortunately, she said that this would not be possible.

It was at this point I realized that even if we managed to finish the collection with the remaining funds, in all likelihood, we would not be able to implement another fix attempt. This was a significant problem. To use the data for the proposed research, it was essential to resolve the systematic error in the database. It could not just be ignored because this group of cohort members who had moved from Uppsala County was likely to be different from those who had stayed. Despite this, several UBCoS steering group members suggested eliminating this part of the database from my proposed project studies.

After an assessment of how much material remained to collect, I also became doubtful about the possibility of successfully completing the entire collection. The additional tasks brought about by the first fix attempt had set the project back by 3 months. I reasoned that my only alternative was to ask my department head for additional funds. My initial and subsequent efforts to secure additional money from the general department fund were rebuffed for almost a year and half.

Irrespective of a clear source of funding, and in anticipation of a future “fix” attempt, I made a decision to track all the errors we found going forward. I reasoned that we would be able to more easily correct them at a later point. I also surmised, correctly in fact, that I would be strongly challenged about the existence of the error problem in the UBCoS database. Tracking the error was a time-consuming process. It had the unfortunate side effect of slowing the remaining archival collection down considerably.

By the start of the summer of 2016, we had gathered more than 90% of the material that we had originally proposed to retrieve. Unfortunately, we were forced to stop due to a shortage of resources. My department had been unresponsive to my ongoing requests for additional support. In the last 3 months of that summer, I made the decision to organize the data we had collected and summarize the additional errors we had seen. I produced a formal report (Garcy, 2016) and presentation with the hope of convincing those in my department that we needed additional funding to finish the collection and address the remaining errors in the database. This effort did not produce the desired effect and I could no longer employ my research personnel.

Ironically, the department decided that it would only fund its own effort to assess what level of error remained in the master file. This effort produced error estimates which were nearly identical to those given in my 2016 report. Ultimately, the first 3-month fix attempt combined with the slower collection schedule added 6 months of unforeseen error correction work to the project.

Practical Lessons Learned

Although I have created a number of databases throughout my career, this was my first attempt to incorporate archival material. It was clearly complicated by several uncontrollable factors. Nevertheless, we nearly succeeded with the archival collection task. One can never fully predict what will happen with such a complex, multiyear collection. However, the following practical strategies and lessons could help to improve your chances of completing a similar archival collection:

- Understand thoroughly the structure of the archives you will visit and the format and availability of the material you seek.
- Consider a pilot survey to help with this task.
- If possible, hire people with archival experience to help with the collection.
- Get personally involved in the collection! Go to the archives! This helps you to develop an overview of problems that may arise and find effective solutions to them. You can also see to it that the collection protocol is adhered to.
- Do not listen to colleagues who tell you that your place is not in the field.
- Have a well-planned collection strategy and schedule.
- Document the material digitally if possible.
- Expect the unexpected, develop contingency plans, and try to maintain some flexibility if events do not go according to plans.
- Back up data on a regular basis! If possible, do it daily—you may not have the opportunity to return to a distant archival location if the data are lost.
- Have a well-planned data entry process. Prepare to be flexible in this process.
- Data collection and entry processes should be done separately. Not at the same time! Database construction mistakes and data entry errors are likely to increase if these processes are combined.
- Do not expect that colleagues will be supportive if problems in an existing database are discovered and not easily solved. Careers and professional reputations can be at stake. Tread carefully!
- Finally, although it is important to be flexible, stick with the collection plan to the best of your ability. Deal with data quality issues after you have all of the data.

Conclusion

This project's data collection was an ambitious and challenging task. I believe that the original goal was attainable within the time and funding constraints of the project. The added task of fixing the corrupted school data in the master UBCoS file proved to be unsurmountable. After the first fix attempt in the summer of 2015, the failure to reduce the error to an acceptable level increased the scope of this task greatly. In retrospect, given the lack of meaningful support from colleagues and my department, the strategy we probably should have pursued was the completion of the collection. If we had ignored the errors until a later time, in all likelihood we would have finished it. Following this strategy may have given us stronger footing for negotiating with the department for additional resources to address all of the problems in the master file. As of the summer of 2017, the remaining errors in the UBCoS database have not been addressed.

Exercises and Discussion Questions

- 1.What would you have done differently to increase the probability of collecting all of the needed material?
- 2.Think about the data entry process. What steps might you take to increase the reliability of the entry process?
- 3.If you were the lead of a project involving a complicated archival data collection, what qualities of the people who you hire to help are important to the success of the project?
- 4.If you had been the PI of this project, would you have tried to fix the database while the collection was ongoing? Give a list of reasons for why you would have refrained from or attempted a fix of this database.

Further Reading

Prior, L. (2003). Using documents in social research. In D. Silverman (Ed.), *Introducing qualitative methods* (pp. 93–110). London, England: SAGE.

Wangard, M. (2017). *Archival research: Using modern techniques to reveal the past*. SAGE Research Methods Cases. Retrieved from <http://methods.sagepub.com/case/archival-research-using-modern-techniques-to-reveal-the-past>

Web Resources

<http://www.chess.su.se/ubcosmg/>

<https://riksarkivet.se/startpage>

References

- Büchele, G., Och, B., Bolte, G., & Weiland, S. K. (2005). Single vs. double data entry. *Epidemiology*, 16, 130–131.
- Garcy, A. M. (2016). UBCoS error report: The current level of error in the corrupted file “UBCoSbetygLisa_SElfrBitte” (No. 1). Report to the UBCoS steering group, Stockholm, Sweden.
- Gibson, D., Harvey, A. J., Everett, V., & Parmar, M. K. B. (1994). Is double data entry necessary? The CHART trials. *Controlled Clinical Trials*, 15, 482–488.
- Gisselmann, M., Koupil, I., & De Stavola, B. L. (2011). The combined influence of parental education and preterm birth on school performance. *Journal of Epidemiology and Community Health*, 65, 764–769.
- Modin, B., Erickson, R., & Vågerö, D. (2012). Intergenerational continuity in school performance: Do grandparents matter? *European Sociological Review*, 29, 858–870.
- Reynolds-Haertle, R. A., & McBride, R. (1992). Single vs. Double data entry in CAST. *Controlled Clinical Trials*, 13, 487–494.
- Van den Berg, G. J., & Modin, B. (2013). Economic conditions at birth, birth weight, ability, and the causal path to cardiovascular mortality (Discussion Paper No. 7605). Bonn, Germany: Institute of Labor Economics.
- Wahi, M. M., Parks, D. V., Skeate, R. C., & Goldin, S. B. (2008). Reducing errors from the electronic transcription of data collected on paper forms: A research data case study. *Journal of the American Medical Informatics Association*, 15, 386–389.
- Wangard, M. (2017). Archival research: Using modern techniques to reveal the past. SAGE Research Methods Cases. Retrieved from <http://methods.sagepub.com/case/archival-research-using-modern-techniques-to-reveal-the-past>
- Williams, A., Garrett, D., & Petroni, R. (2004). Quality control of data entry for the American Community survey and the impact of errors on data quality. In Paper presented at Joint Statistical Meetings (pp. 423–425). Toronto, Ontario, Canada: American Statistical Association Section on Survey Research Methods.