

Record Linkage and Public Policy – A Dynamic Evolution

Ivan P. Fellegi, Statistics Canada

Abstract

Record linkage, as a major domain of substantive and technical interest, came about in the 1960s at the confluence of four closely inter-related developments:

- *First, the post-war evolution of the welfare state and taxation system resulted in the development of large files about individuals and business.*
- *Second, new computer technology facilitated the maintenance of these files, the practically unlimited integration of additional information, and the extraction of hitherto unimaginably complex information from them.*
- *Third, the very large expansion of the role of government resulted in an unprecedented increase in the demand for detailed information which, it was thought, could at least partially be satisfied by information derived from the administrative files which came about precisely because of this increase in the role of government.*
- *But there was a fourth factor present in many countries, one with perhaps the largest impact on subsequent developments: a high level of public concern that the other three developments represented a major threat to individual privacy and that this threat had to be dealt with.*

The paper traces the dynamic interaction of these factors over time, and their impact on the evolution of record linkage practice in three different domains of application: government statistics, public administration, and the private sector.

Introduction

It is a great honour and pleasure to be here today to think out loud about record linkage. I will say a few words about its evolution, its current status, and I will share with you some reflections about the future.

Let me start with the simplest possible definition of record linkage. There is a single record as well as a file of records and all records relate to some entities: persons, businesses, addresses, etc. Record linkage is the operation that, using the identifying information contained in the single record, seeks another record in the file referring to the same entity. If one accepts this definition, it is clear that people have been linking records ever since files existed: the filing clerk, for example, spent his or her entire working day looking for the “right” file to retrieve, or to insert new material in. The “right” file, of course, was the one that corresponded to the identification that was sought, where “identification” could be anything that uniquely described the “right” file.

Of course, this description of the traditional record linkage appears to be circuitous, but this did not matter to anyone: everyone knew what needed to be done, so the lack of definition had no operational consequence. The human mind could recognize the identification of a record in a file -- whether or not the descriptors contained some errors.

Four Critical Factors Shaping the Evolution of Record Linkage

Record linkage, as a major domain of substantive and technical interest, came about in the 1960s at the confluence of four closely inter-related developments:

- First, the post-war evolution of the welfare state and taxation system resulted in the development of large files about individuals and businesses (**opportunity**).
- Second, new computer technology facilitated the maintenance of these files, the practically unlimited integration of additional information, and the extraction of hitherto unimaginably complex information from them (**means**).
- Third, the very large expansion of the role of government resulted in an unprecedented increase in the demand for detailed information which, it was thought, could at least partially be satisfied by information derived from the administrative files which came about precisely as a consequence of the increase in the role of government (**need**).
- But there was a fourth factor present in many countries, one with perhaps the largest impact on subsequent developments: a high level of public concern that the other three developments represented a major threat to individual privacy and that this threat had to be dealt with (**constraint**).

As a result of the fourth factor, there was a very real commitment in these countries, whether formally taken or only implicitly accepted, that the creation of population registers must be avoided, indeed that even a uniform system of identifying persons would be unacceptable. In effect, files would be set up as and when needed, but personal information would not be integrated in a comprehensive manner. If all the relevant information had been kept together in a single large register, there would clearly have been little motivation to carry out the complex task of bringing together information from large and distinct files that were not designed for the purpose. Record linkage, it is important to remember, was therefore of particular interest in those countries which had a long history of striking a balance in favour of the individual in the tension between individual rights versus the needs of the state. In effect, record linkage came about to accomplish a task that was rendered difficult precisely because there was a social consensus that it **should** be difficult.

Much of the paper is devoted to an exploration of the dynamics among these four factors, and how they played themselves out in different domains of application: the statistical domain, other government applications, and the private sector. For simplicity and focus, I will mostly restrict my comments to files involving personal information.

A Historical Digression

While record linkage flourished because of government's administrative need, it is important to remember that the pioneering work of Howard Newcombe (Newcombe and Kennedy, 1962) was motivated by interest in genetic and biomedical research. Indeed, to this day a majority of record linkage applications carried out in Statistics Canada are health related.

However, my work with Alan Sunter (Fellegi and Sunter, 1969) was not motivated by health research issues. Rather, it was explicitly oriented to the problem of merging the information content of large administrative files in order to create a statistically useful source of new information. Our contribution can be summarized as follows:

- Newcombe recognized that linkage is a statistical problem: in the presence of errors of identifying information to decide which record pair of potential comparisons should be regarded as linked. Our first contribution involved formalizing this intuitive recognition and rigorously describing the space of record pairs consisting of all possible comparisons;
- Second, we provided a calculus for comparing the evidence contained in different record pairs about the likelihood that they refer to the same underlying unit;
- Third, we defined a linkage rule as a partitioning of a comparison space into the subset that we called “linked,” i.e., record pairs about which the inference is that they indeed refer to the same underlying unit, a second subset for which the inference is that the record pairs refer to different underlying units, and a complementary third set where the inference cannot be made without further evidence;
- Fourth, as a formalization of the statistical character of the linkage rule, we identified the characteristic Type I and Type II errors associated with a given linkage rule: the proportion of record pairs that are falsely linked and the proportion that are incorrectly unlinked;
- Fifth, we showed that if the space of record pair comparisons is ordered according to our metric, this will result in a linkage rule that is optimal for any pre-specified Type I and Type II error levels;
- Our final contribution was to provide a framework that, in retrospect, turned out to be fruitful both for the design of operationally efficient record linkage systems and for the identification of useful areas for further research. Perhaps this was our most important contribution: facilitating the outstanding research that followed.

The very fact that there is a successful symposium here today is testimony to the productivity of that research. There has certainly been a spectacular evolution of methodology and techniques, signalling a continuing, perhaps even increasing interest in the topic. However, the basic tension among the four critical factors mentioned above was never fully resolved, even though the dynamics were quite different in the different domains of application. Let me turn to a brief overview of these application domains.

The Statistical Domain -- A Model?

Since most of us here are statisticians, I will start with statistical applications. The defining characteristic of this domain is that the output does not relate to identifiable individuals -- i.e., that statistical confidentiality is preserved. This very important distinction ought to result in a different public attitude to linkage by statisticians for statistical purposes. But I am not sure that it does -- for two related reasons. First, the process of linkage of personal records is intrinsically privacy intrusive, in the sense that information is brought together about a person without his or her knowledge and control. From that point of view it is largely irrelevant that the results can only affect particular individuals in an indirect manner. The second reason is that not everyone trusts us completely to maintain statistical confidentiality.

Statistical confidentiality protection in Canada, at least within government, is certainly tight -- both legally and de facto. As you know, unlike the United States, almost all government statistical activity is carried out within a single agency and is covered by a uniform and strong statistics act. In spite of that, we have taken what we think is a very cautious, though hopefully balanced attitude to record linkage. We have developed explicit

policies and strong mechanisms to make them effective.

Statistics Canada will undertake record linkage activities only if all the following conditions are satisfied:

- the purpose of the linkage activity is statistical/research;
- the products of the activity will be released only in accordance with the confidentiality provisions of the Statistics Act;
- the benefits to be derived from the linkage are substantial and clearly serve the public interest;
- record linkage is either the only option to acquire the needed information or, given the cost or burden implications of alternative approaches, it is the only feasible option;
- the record linkage activity will not be used for purposes that can be detrimental to the individuals involved;
- the record linkage activity is judged not to jeopardize the future conduct of Statistics Canada's programs; and finally
- the linkage satisfies a prescribed review and approval process.

Let me underline some features of this policy. Beyond the more or less obvious fact that we will not carry out linkage except for statistical purposes, and that we will protect confidentiality, the main feature of the policy is to seek a balance. We recognize that linkage is intrinsically intrusive of privacy, so we will only consider undertaking it where the public benefit is sufficiently important to tip the balance of decision. But even when this is the case, we want evidence that alternative methods to acquire equivalent information are either impossible or prohibitively costly. Another requirement is that the objective of the project should not be detrimental to the individuals concerned. This last point bears emphasis: we are not talking about individual jeopardy -- that is ruled out by our strict confidentiality protection -- but of possible harm to the group of people whose records are involved. Since typically we cannot contact them to obtain their informed consent, we want to make sure that we will not link their information if, as a group, they would not have given us informed consent had we been able to seek it.

However noble, no set of principles is likely to have operational impacts without a set of procedures designed to give them effect. In our case this involves a cascading set of approvals. Every manager who wishes to sponsor a linkage application has to submit a narrative describing the purpose, the expected public benefits, whether there is a possibility of harm to the individuals concerned, and whether there are feasible alternative approaches. In addition, the manager also has to describe the proposed methodology, any features that might enhance privacy or confidentiality protection, and has to propose a tight schedule for the destruction of linked identifiers. This information is assessed, in the first place, by a standing committee composed of several of Statistics Canada directors that is chaired by one of my direct assistants. Their assessment and recommendation is reviewed by the agency's top level management group that I chair. If we decide that the public good indeed outweighs the privacy concern and that the objectives cannot reasonably be achieved without linkage, we next consider whether the project needs ministerial approval and/or external "stakeholder" review. Generally, ministerial approval is sought unless a previous approval has clearly established a precedent.

Least problematic are cases where both files to be linked were collected for a statistical purpose. Examples are routine linkages of successive rounds of panel surveys for purposes of editing, linkage of successive waves of longitudinal surveys, or the linkage of the census and a post-censal survey to assess the completeness of the census count.

Health applications provide another class of well established precedents. These typically involve a file, provided by an external organization, containing records of persons known to have been exposed to a health risk. These exposure files are linked either to a machine readable cancer register or to our mortality file. The purpose is to assess whether the exposed persons had a higher rate of some specified cancer, or whether they had a disproportionate number of deaths due to a suspected cause. If the proposal involves more than a scientific fishing expedition, i.e., if it is designed to explore a reasonably well-founded scientific hypothesis, then the linkage is normally approved by the senior management of Statistics Canada, since the precedent for ministerial approval is well established in these cases.

In other cases where there is no applicable precedent, the public benefit is considered carefully by Statistics Canada. If in our judgement the benefit is not sufficient to proceed, the request is rejected and that is the end of the matter. However, if we feel that there is considerable public benefit to be derived from the linkage and there is no alternative approach that is practical, then we make a positive recommendation to our minister.

Statistics Canada is a very autonomous organization, operating at arm's length from the political process. This is just about the only programmatic issue on which we seek ministerial guidance. Why do we do so here? Because privacy and information are both public goods and there is no methodology for a **professional** assessment of the right balance between them. Establishing the balance between competing public goods, however, is very much a function of elected politicians in democratic societies.

Our review process might involve an extra step in those rare cases where the potential public good is judged to be very high, but where the privacy issue is particularly sensitive. An example will illustrate. In Canada a substantial proportion of the social assistance to the poor is administered by provinces and there is no federal record of such disbursements. Conversely, the provinces have no access to the records of federal social assistance programs, e.g., unemployment insurance. A few years ago we received a very serious research proposal to study the combined effect on incomes of federal and provincial social assistance, as well as taxes. The objective was to assess the combined impact of all these programs: e.g. are they properly focused on the poor, or are there unintended disincentives to work. This was clearly a program of major potential public benefit, indeed of major potential benefit to the poor. But, equally clearly, it would have been prohibitively expensive to secure their informed consent. As the next best thing, we convened a seminar with the Privacy Commissioner, with interested researchers, and with representatives of advocacy groups for the poor, calling on the latter as proxy spokespersons. The linkage was endorsed by them and it subsequently received ministerial approval.

Our approach to privacy protection merited the following salutation from the Privacy Commissioner in his annual report to Parliament in 1990:

“Worthy of special mention in an end-of-term report is the tangible commitment to privacy demonstrated by the Chief Statistician of Canada. It is especially noteworthy because many of the Privacy Act requirements do not apply to statistical records.

The Chief Statistician took the initiative, as he has on other privacy matters, and sought the Privacy Commissioner's view on whether the public interest justified conducting this record linkage.

The Privacy Commissioner agreed that the proposed pilot project had potential for contributing significantly to the public interest; most important, he considered it impossible to accomplish the goal without intruding on personal privacy...

Some may consider the Privacy Act remiss in not subjecting personal information used for statistical purposes to the same requirements imposed on personal information used for administrative purposes. No one should doubt, however, that the privacy concerns about statistical data are being addressed in practice.

The Chief Statistician of Canada is to be thanked for that.”

I have quoted from this report at some length because it makes a number of important points. First, even a professional privacy advocate like the Commissioner agrees explicitly that the need to protect privacy must be weighed against the need for information. Second, because it makes clear that while statistical records are not legally subject to most requirements of the Privacy Act, it is strategically important for statisticians to be very prudent about record linkage. And last but not least, the quote (and our continuing experience since 1990 as well) proves that a well balanced policy, **together with concrete administrative practice to give it effect**, can effectively mediate between the two competing public goods of privacy and need to know.

The Dog That Does Not Bark: Public Administration and Public Unconcern

Large government data banks about persons have traditionally been regarded as threatening because of the visible power of the state: to make compulsory the provision of information that it needs, to make decisions on the basis of information in its possession, and finally to enforce the decisions made on the basis of the information held. One might expect, therefore, that the balance of forces affecting record linkage would follow Newton’s third law of dynamics: the stronger the combined pro-linkage forces of opportunity, need and means, the stronger would become the constraining counterforce of concerns about privacy. Yet, during much of the last thirty years there has been a curious disjunction between, on the one hand, the level of public anxiety about record linkage and data banks, and on the other hand the level of actual government activity in these fields.

During the 1960s, 70s and much of the 80s, two main factors restrained the spread of government record linkage applications. First, the potential pay-off from this kind of linkage -- reduction of financial errors and outright fraud -- was not high on the agenda of governments. Consequently the widespread, though mostly latent, public hostility toward linkage effectively restrained governments, particularly in the absence of an overriding financial objective. And second, the level of technology available at the time kept the cost of linkage reasonably high. In addition, some measures of transparency introduced in most developed countries have also been helpful in alleviating concerns. In Canada, for example, every citizen has access to a register which describes the content and purposes of all government held personal data banks. Should they wish, they may obtain, free of charge, a copy of the information held about them. They also have a legal right to have their non-statistical records corrected or updated if they are in error.

The combination of these factors has effectively blunted the public’s sense of concern about government data banks and record linkage. Yet during the last several years a significant change occurred in the balance of the forces at work. Deficit reduction rose to the top of governments’ agenda, increasing the importance attached to controlling tax evasion and welfare fraud. At the same time, cuts in public services in the name of deficit reduction have caused substantial and widely reported hardships -- compared to which privacy fears seemed to have assumed a diminished importance. And, of course, the cost of linkage shrank rapidly. So, precisely at a time when record linkage applications by government are growing rapidly, there is hardly any public debate, let alone open concern, about the practice -- except for the one-day news triggered annually by the release of yet another report by the Privacy Commissioner.

So long as the public is properly informed but chooses to be unconcerned, we should all be pleased about the equilibrium that may have been reached. But I am concerned. First of all because the apparent equilibrium is not based on informed debates. Even more important, I believe the status quo to be fragile: a single egregious

error or accident might suffice to put a spotlight on the extent of linkage going on in government. In that case the incident might well balloon out of control if elementary questions cannot be answered about the weight given to privacy issues in approving each application, about procedural checks and balances, about accountability, and about the point in the bureaucratic and political hierarchy at which final approval was given.

I believe there is a great need for much increased transparency here. **We need to develop explicit and publicly debatable policies about both the criteria and processes involved in approving record linkage for administrative purposes** -- perhaps along the lines of the process used by Statistics Canada, but of course suitably modified to fit the different domains of application. This need not go as far as it has in some European countries where approval depends on an appointed privacy commissioner. Privacy Commissioners have an important role as advocates, i.e., as the public guardians of one side of the issue. But approval should entail a proper consideration of the conflict between the two competing public goods: privacy and enforcement. As such the approval should ultimately involve the political level. The process of political consideration can and must be supported by a bureaucratic process which reviews options, assesses benefits, and recommends approaches that reduce the risks to privacy.

Out of Control -- Linkage in the Private Sector

My concerns about the public sector are dwarfed by the discomfort I have about linkage in the private sector. Not that I know much about it -- and I suspect the same applies to most of you. But this is precisely the sign of a potentially very serious problem: the unrecognized and undiscussed threat of privately held data banks and large scale record linkage.

On the surface, and in comparison with the public sector, the private sector appears to be innocuous for two broad and interconnected reasons. First, one may think that its possible decisions about us affect us less profoundly, and hence the issue of control over personal information is less acute. And second, that the private sector has no legally enforceable sanctions to back up its data collection and its decisions about people. But none of the arguments about lack of sanctions or lack of impact stand up to scrutiny. On the one hand, the unregulated private sector rarely has the need to back up its information based decisions with sanctions against individuals -- it simply stops dealing with them. On the other hand, the ultimate threat of denying a benefit is probably sufficient to make the collection of information compulsory to all intents and purposes. After all, try to obtain a credit card, register a warranty, or seek insurance without providing the requested information. Of course, it can be argued that having a credit card or health insurance are not necessities -- but would you like to try living without them?

While some segments in the private sector clearly have de facto compulsory powers of data collection, others completely bypass the issue of informed consent by buying personal information. The impact of decisions on our lives made on the basis of indirectly obtained information can range from the inconvenient to the profound. Let me give you some examples.

You may or may not receive certain kinds of advertising depending on the information held about you by the distributor. You may miss some information that you might like to have or, conversely, you might be annoyed by the unnecessary "junk mail."

The information held about you might affect your credit rating without you even being aware, let alone having the power to insist on the correction of erroneous information.

You may not receive insurance you might like to receive.

Your adversaries in a court case might gain undue advantage in preparing their case by accessing information held about you.

The list of examples could go on. I hope to have convinced you that there is, indeed, a serious privacy risk. How come that, as a society, we have allowed this situation to evolve? In effect, the four factors affecting record linkage have evolved differently in this domain compared to others.

- **Opportunity.**--The availability of cheap and powerful hardware and software facilitated the widespread use of information and communication technology. Therefore a capacity was created and made widely accessible for building up large files about clients as well as for making use of similar files created by others;
- **Means.**--Fragmented and dispersed lists of persons, even if widely available, used to be regarded as representing a negligible risk. But the cheap availability of computing, together with the powerful methodology-based software has altered the picture. Indeed, we in Statistics Canada were able to construct, as part of our 1991 census preparation, an excellent address list using client lists from the tax authority, telephone companies, hydro companies and municipal assessments. Even more worrisome is that the 3-4 largest credit card companies, among themselves, hold a list of consumers whose coverage of the adult population is probably close to universal. Furthermore, in addition to their excellent coverage, credit card company lists also hold a vast amount of information about us, including our income, expenditure patterns (and therefore our individual preferences and dislikes), our travel pattern, home repairs, etc.
- **Need.**--Advertising has evolved and it no longer relies solely on the mass media. Increasingly, companies prefer pinpoint approaches, using a variety of mailing or telephone lists. These lists are customized with great sophistication, using the information contained in them, to delineate the population group to be targeted. The advertising utility and value of the input files is directly determined by the amount of relevant personal information contained in them.

Thus three of our four critical factors have interacted positively, resulting in the creation of a mass market for large electronic files about consumers. In turn, this mass market created a new industry of service providers. These are information brokers specializing in consumer files, their updating, the upgrading of information held about people (i.e., record linkage), and the marketing of both the files themselves and services based on them. Although I have no objective evidence, there is little doubt that commerce in personal information has become a big business. And it is entirely unregulated. Which leads to the fourth critical factor:

- **Constraint.**--As indicated above, there is practically none.

From the perspective of privacy there are two quite distinct problems with the current situation. The first is that we have, indeed, lost control over the use made by others of information held about ourselves. The second basic problem is that we can't even control the accuracy of this information.

So there is a problem. Is there a solution?

The knee-jerk reaction might be to address these problems through regulation. But it only takes a few seconds of reflection to realize that this traditional tool, by itself, would not be workable. Electronic communication has become so cheap that the physical location of files can be moved anywhere in the world without the least impact on access and use. So if they don't like one country's regulations, the information brokers can simply take their files to another.

Are we completely defenceless? I don't think so. If regulations cannot be enforced by government, perhaps they can be designed to lead to a degree of self-enforcement within the private sector, based on their own enlightened self-interest. I will conclude this talk by outlining a possible approach.

Ideally, one may wish to achieve two objectives: to improve people's control over the use of information

about themselves; and to improve their ability to control the accuracy of such information.

I am not particularly sanguine about restoring individuals' control of the use of information about themselves. But we might nevertheless be able to improve the current situation. The approach could be based on the observation that while the operations of companies might be moved from one country to another, the transaction whereby members of a population provide information about themselves is intrinsically a domestic one, hence potentially subject to regulation. We could prescribe, therefore, that when information is requested from people, certain information must also be provided to them. This could include a description of the information management practices of the requesting company, the control methods they use, and whether and under what conditions they provide access to personal information to other companies. Can the government effectively monitor the adherence by companies to such standards? Certainly not directly. But at this point competitive pressures might come to the fore. Those who adhere to their promised information management standards incur some costs. It is in their interest that their competitors' misdemeanours should not be a source of unfair competitive advantage to the latter. The availability of formal complaint mechanisms, maintained either by the government or by the industry itself, might provide a constructive outlet for the policing of each firm's practice by its own competitors.

This approach would not be a guarantee against undesirable secondary information use. But at least it might go some distance towards a form of informed consent. If there is a significant number of consumers for whom control of their information is a priority, competitive pressure might help encourage some firms to cater to such consumer preferences, even at some small additional cost. If, as a result, these firms end up with better information, they will gain market share, eventually, perhaps squeezing out their less accommodating competitors. But in order to encourage this form of competition, government must provide a productive framework through its initial regulation and the creation of a suitable complaint mechanism.

I am a little more optimistic regarding the possibility of improving the reliability of the information content of personal information banks held in the private sector. First, accuracy of information is surely in the interest of an overwhelming proportion of users of personal information. So government could establish a licensing system for recognized carriers of personal information. The data banks held by such carriers (whether or not they are physically inside or outside the country) would be listed in a register of personal data banks, such as exists in Canada in respect of government operated personal data banks. A requirement for the license would be the obligation to provide free access by people to the information held about themselves, as well as the obligation to implement all corrections on demand. Since compliance with these regulations would improve the accuracy of information held in such data banks, registered carriers would have a competitive advantage. However, adherence to the regulations would also drive up their costs. If the competitive advantage due to higher accuracy is not sufficient to offset their higher costs, it might be necessary to consider additional measures. One possibility would be penalties against the usage of personal information held in unregulated data banks (the penalties would, of course, have to be assessed against the information users rather than the providers since the latter might be outside the national territory). The price differential between the regulated and unregulated operators would provide an incentive for the former to "police" the latter.

The combination of approaches proposed here would not restore to individuals full control over information about themselves -- the ultimate objective of privacy advocates. But I am firmly convinced that this objective is no longer attainable. They would, however, restore a significant element of informed consent to the process of providing personal information. It would also go some way to improving the accuracy of privately held personal information

banks. As such, these measures would be a major improvement over the current absolute free-for-all -- a situation which, I believe, is intolerable. If we do not at least acknowledge that we have a serious and rapidly worsening problem, and if we do not take practical measures to deal with it, then in effect we connive in its continuation and exacerbation.

Conclusion

We seem to have come full circle. As a society we did not want comprehensive population registers, largely because we did not want a large scale and routine merging of information contained in government files. But we did not want to rule out *some* merging for *some* well justified purposes. So, as a matter of conscious public policy, we made linkage very difficult. However, we allowed the development of record linkage methodology for use in exceptional circumstances. The applications were indeed important, often requiring a high level of accuracy, so we refined the methodology, and also made it vastly more efficient. Combined with rapidly diminishing computing costs, this efficient methodology rendered linkage into a truly ubiquitous tool: indeed, at this symposium there are several versions of the methodology on display, competing on efficiency and ease of use. The activity that was designed to be difficult has become quite easy and inexpensive.

As a society we have been concerned with the power of the state and the risk of that power being abused. So we constrained the ability of the state to use our personal information without our consent. It is perhaps a coincidence, but certainly not a contradiction, that as the relative power of the state is declining, we are, *de facto* and without much public discussion, allowing it more extensive latitude to link our personal information without our explicit agreement. It is, however, a paradox that as the relative balance of power is shifting to the private sector, we are allowing it to build up extensive personal data banks, without regulation or even assurances about the accuracy of its contents. The power that we used to be anxious to deny to the state, which is operating under the guidance of our elected representatives, we are allowing the private sector to acquire -- indeed we seem to be doing so with a shrug of the shoulders.

In a democratic society it is of paramount importance that major public issues be decided based on well informed public debate. This paper is intended as a modest contribution to ensuring that our **consent** is, indeed, based on **understanding**.

References

Fellegi, I. P. and Sunter, A. B. (1969). A Theory for Record Linkage, *Journal of the American Statistical Association*, 64, 1183-1210.

Newcombe, H. B. and Kennedy, J. M. (1962). Record Linkage: Making Maximum Use of the Discriminating Power of Identifying Information, *Communications of the A.C.M.*, 5, 563.