The frequency of misattributed paternity in Sweden is low and decreasing: A nationwide cohort study

Torsten Dahlén1,2, Jingcheng Zhao1, Patrik K.E. Magnusson3, Yudi Pawitan3, Jakob Lavröd4 & Gustaf Edgren1,5

From the 1Department of Medicine Solna, Clinical Epidemiology Division, Karolinska Institutet, Stockholm, Sweden; 2Department of Hematology, Karolinska University Hospital, Stockholm, Sweden; 3Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 4Department of Mathematics, Lund University, Lund, Sweden; and 5Department of Cardiology, Södersjukhuset, Stockholm, Sweden

Abstract. Dahlén T, Zhao J, Magnusson PKE, Pawitan Y, Lavröd J, Edgren G. The frequency of misattributed paternity in Sweden is low and decreasing: A nationwide cohort study. J Intern Med. 2022;291:95-100.

Background. The occurrence of misattributed paternity has consequences throughout society with implications ranging from inheritance and royal succession to transplantation. However, its frequency in Sweden is unknown.

Objective. To estimate the contemporary frequency of misattributed paternity in Sweden.

Methods. The study was based on nationwide ABO blood group data and a nationwide register of familial relationships in Sweden. These data were analysed using both a frequentist Poisson model and the Bayesian Gibbs model. The conduct of the study was approved by the regional ethics committee in Stockholm, Sweden (reference numbers 2018/167-31 and 2019-04656).

Results. Nearly two million mother-father-offspring family units were included. Overall, the frequency of misattributed paternity was estimated at 1.7% in both models. Misattributed paternity was more common among parents with low educational levels, and has decreased over time to a current 1%.

Conclusions. The misattributed paternity rate is similar to the rates in other West European populations. Apart from widespread societal implications, studies on heritability may consider misattributed paternity as a minor source of error.

Keywords: extra-pair paternity, false paternity, misattributed paternity, multi-generation register, SCANDAT

Introduction

Misattributed paternity, where a man is incorrectly believed to be the genetic father of a child, has widespread implications ranging from sibling bone marrow transplantations to inheritance disputes, and royal succession orders [1]. Given its sensitive nature, it has been difficult to estimate its true frequency and no population-based estimates are available for the Swedish population [2].

Several back-dated studies, presenting case reports or investigating the frequency of misattributed paternity, have based their analyses on non-representative populations, including men seeking paternity testing, patients undergoing compatibility testing for organ transplantation, individuals seeking genetic counselling or in genealogically well-enumerated populations [3–9]. Thus, differences in estimates of the misattributed paternity occurrence can likely be attributed not only to population variability but also to variations in pre-test probabilities. From more recent studies, estimating the misattributed paternity rate over several generations using more sensitive genomic methods, the rate seems to be considerably lower than early estimates [10]. For example, in a study combining Y-chromosome genotypes with genealogical data in a Flemish population, the frequency of misattributed paternity was estimated at 1–2% per generation during the last 400 years [11]. Contemporary rates, in
The ABO blood group system is an autosomal triallelic system located at a single locus, with alleles denoted as A, B and O. The blood group phenotypes and accompanying allele combinations are A (AA or AO), B (BB or BO), AB (AB) and O (OO), rendering A and B co-dominant with the autosomal recessive inheritance of O. The allele frequencies are denoted as p, q and r for A, B and O, respectively. As such, impossible inheritance patterns can be determined depending on the mother’s and reported father’s blood group. For all analyses, we estimated blood group allele frequencies using the blood group distribution from the SCANDAT3-S database, assuming Hardy-Weinberg equilibrium. To estimate the allele frequencies, p, q and r, in our population, we used the expectation-maximization algorithm with six iterations [19].

Statistical analyses

We used a two-pronged approach to estimate the frequency of misattributed paternity. First, where the frequency of misattributed paternity was estimated using the occurrence of impossible blood group combinations (i.e. where the offspring had a blood group that would not be possible given the parents’ blood group combination, such as when both parents had blood group O and the offspring had a non-O blood group). The analyses considered the population blood group allele frequencies in order to account for the possibility that the genetic father may have the same blood group as the attributed father. Observed impossible and possible blood group combinations are demonstrated in Supporting Information Fig. S1. For every impossible blood group combination, we calculated an expected count derived from the mother’s allele frequency and the possible allele frequency of the genetic father. The Poisson regression model was specified based on the 16 impossible mother–father offspring relations depicted in Supporting Information Table S1. We assume that the frequency of misattributed paternity is independent of the blood groups of the parents and that the frequency is sufficiently low to be approximately Poisson distributed. We fit this model as a Poisson regression model using a log link function, with the total number of family units per mother, father and offspring blood group combination as the outcome. The logarithm of the product of the number of impossible parents and known mother and genetic father expected genetic contribution was used as an offset term (Supporting Information Table S1), yielding the logarithm of the probability of misattributed paternity as the intercept. The model was tested for equidispersion. We constructed 95% confidence intervals using the Wald method.

Methods

Data sources

We combined all available ABO blood group data from the transfusion database SCANDAT3-S (Scandinavian Donations and Transfusions) [17] with family structure data from the Swedish multigeneration register recording family information for all individuals born in 1932 and alive in 1954. Most individuals born between 1947 and onwards were included but information on individuals who immigrated was missing until 1997 [18]. Blood donors were removed from the database as they were deemed to be non-representative of the general population. The resulting database thus contained data on families where both parents and at least one offspring had undergone ABO blood group testing in preparation for surgery, prior to a transfusion, or for routine testing in antenatal care or at childbirth. Family units were constructed with information on mother, father and offspring blood group. As such, families with multiple offspring were recorded as separate family units. Information on the educational level was retrieved from Statistics Sweden, the governmental agency responsible for population statistics. Further details about the study design and some baseline characteristics of the cohorts are found in Fig. S1.

ABO blood group system

The ABO blood group system is an autosomal triallelic system located at a single locus, with alleles denoted as A, B and O. The blood group phenotypes and accompanying allele combinations are A (AA or AO), B (BB or BO), AB (AB) and O (OO), rendering A and B co-dominant with the autosomal recessive inheritance of O. The allele frequencies are denoted as p, q and r for A, B and O, respectively. As such, impossible inheritance patterns can be determined depending on the mother’s and reported father’s blood group. For all analyses, we estimated blood group allele frequencies using the blood group distribution from the SCANDAT3-S database, assuming Hardy-Weinberg equilibrium. To estimate the allele frequencies, p, q and r, in our population, we used the expectation-maximization algorithm with six iterations [19].

Statistical analyses

We used a two-pronged approach to estimate the frequency of misattributed paternity. First, where the frequency of misattributed paternity was estimated using the occurrence of impossible blood group combinations (i.e. where the offspring had a blood group that would not be possible given the parents’ blood group combination, such as when both parents had blood group O and the offspring had a non-O blood group). The analyses considered the population blood group allele frequencies in order to account for the possibility that the genetic father may have the same blood group as the attributed father. Observed impossible and possible blood group combinations are demonstrated in Supporting Information Fig. S1. For every impossible blood group combination, we calculated an expected count derived from the mother’s allele frequency and the possible allele frequency of the genetic father. The Poisson regression model was specified based on the 16 impossible mother–father offspring relations depicted in Supporting Information Table S1. We assume that the frequency of misattributed paternity is independent of the blood groups of the parents and that the frequency is sufficiently low to be approximately Poisson distributed. We fit this model as a Poisson regression model using a log link function, with the total number of family units per mother, father and offspring blood group combination as the outcome. The logarithm of the product of the number of impossible parents and known mother and genetic father expected genetic contribution was used as an offset term (Supporting Information Table S1), yielding the logarithm of the probability of misattributed paternity as the intercept. The model was tested for equidispersion. We constructed 95% confidence intervals using the Wald method.

Methods

Data sources

We combined all available ABO blood group data from the transfusion database SCANDAT3-S (Scandinavian Donations and Transfusions) [17] with family structure data from the Swedish multigeneration register recording family information for all individuals born in 1932 and alive in 1954. Most individuals born between 1947 and onwards were included but information on individuals who immigrated was missing until 1997 [18]. Blood donors were removed from the database as they were deemed to be non-representative of the general population. The resulting database thus contained data on families where both parents and at least one offspring had undergone ABO blood group testing in preparation for surgery, prior to a transfusion, or for routine testing in antenatal care or at childbirth. Family units were constructed with information on mother, father and offspring blood group. As such, families with multiple offspring were recorded as separate family units. Information on the educational level was retrieved from Statistics Sweden, the governmental agency responsible for population statistics. Further details about the study design and some baseline characteristics of the cohorts are found in Fig. S1.

ABO blood group system

The ABO blood group system is an autosomal triallelic system located at a single locus, with alleles denoted as A, B and O. The blood group phenotypes and accompanying allele combinations are A (AA or AO), B (BB or BO), AB (AB) and O (OO), rendering A and B co-dominant with the autosomal recessive inheritance of O. The allele frequencies are denoted as p, q and r for A, B and O, respectively. As such, impossible inheritance patterns can be determined depending on the mother’s and reported father’s blood group. For all analyses, we estimated blood group allele frequencies using the blood group distribution from the SCANDAT3-S database, assuming Hardy-Weinberg equilibrium. To estimate the allele frequencies, p, q and r, in our population, we used the expectation-maximization algorithm with six iterations [19].

Statistical analyses

We used a two-pronged approach to estimate the frequency of misattributed paternity. First, where the frequency of misattributed paternity was estimated using the occurrence of impossible blood group combinations (i.e. where the offspring had a blood group that would not be possible given the parents’ blood group combination, such as when both parents had blood group O and the offspring had a non-O blood group). The analyses considered the population blood group allele frequencies in order to account for the possibility that the genetic father may have the same blood group as the attributed father. Observed impossible and possible blood group combinations are demonstrated in Supporting Information Fig. S1. For every impossible blood group combination, we calculated an expected count derived from the mother’s allele frequency and the possible allele frequency of the genetic father. The Poisson regression model was specified based on the 16 impossible mother–father offspring relations depicted in Supporting Information Table S1. We assume that the frequency of misattributed paternity is independent of the blood groups of the parents and that the frequency is sufficiently low to be approximately Poisson distributed. We fit this model as a Poisson regression model using a log link function, with the total number of family units per mother, father and offspring blood group combination as the outcome. The logarithm of the product of the number of impossible parents and known mother and genetic father expected genetic contribution was used as an offset term (Supporting Information Table S1), yielding the logarithm of the probability of misattributed paternity as the intercept. The model was tested for equidispersion. We constructed 95% confidence intervals using the Wald method.
Secondly, we also fitted a Bayesian model that was based on all mother-father-offspring combinations, both possible and impossible combinations. The probability for each combination to have a non-genetic father was derived conditionally on the frequency of misattributed paternity, by demonstrating that the probability of a specific phenotype in the offspring is the sum of the conditional probabilities of offspring of a specific phenotype given to mother and non-genetic father and mother and genetic father, weighted by the probability of misattributed paternity. The calculations were carried out using the Markov Chain Gibbs sampling scheme [20]. For each step in the chain, the misattributed paternity from the previous step in the chain was used to compute the conditional probability for each combination of the family’s blood type, allowing the number of genetic fathers to be sampled. When the number of families in a specific combination was greater than 25, a Poisson approximation to the binomial distribution was used to speed up the computations. The uncertainty of the blood group distribution, and hence allele frequency distributions due to limited data for a specific cohort, was included in the sampling for each iteration. However, the effect sizes or uncertainty intervals did not change owing to a large number of individuals in the cohorts. Once the number of genetic fathers out of the total number of families had been obtained in a specific step in the chain, this number was used to sample the misattributed paternity fraction using a beta distribution, with \( \beta(2,2) \) prior to excluding the case of certainty concerning the parental status, which in turn was used to initiate the calculations in the next iteration of the Gibbs chain. After discarding \( 10^5 \) burn-in chain iterations, an additional \( 10^6 \) iterations were completed to obtain the estimates and accompanying uncertainty intervals.

Using the second model, we also performed analyses restricted to children born before in vitro fertilization was introduced in Sweden in 1983, as well as stratified by a decade of birth and parental education. As online-only supplements, we also provide estimated rates for analysis on the geographical region of birth and maternal and paternal age at birth (Supporting Information Figs S2 and S3).

Data processing and statistical analyses were conducted using SAS Statistical Analysis Software, version 9.4, R, version 3.6.0 or MATLAB2019a, as appropriate.

Results

A total of 1.95 million family units were included in the analyses. The blood group distribution in the full study population was 45% A, 5% AB, 12% B and 38% O. The cohort mainly consisted of offspring born between 1950 and 1990. Of all parents, 89% were born in Sweden and of the offspring, 97% were born in Sweden.

The main analyses resulted in estimates of the frequency of misattributed paternity that were very similar between the Poisson, 1.7% (95% confidence interval, 1.6–1.7%) and Bayesian models, 1.7% (95% credibility interval, 1.6–1.7%) (Fig. 1a). In a sensitivity analysis, to address the issue of non-independent samples, we restricted the cohort to the first child born in every family unit. The estimates were similar and again consistent between the Poisson and Bayesian models, 1.8 (95% confidence interval, 1.7–1.9%) and 1.8 (95% credibility interval, 1.7–1.8%). When the analyses using the Bayesian model were restricted to offspring born before in vitro fertilization, first introduced in Sweden in 1983, it yielded a slightly higher estimate, 2.0 (95% credibility interval, 1.9—2.0%). In turn, this was consistent with a general decline in the occurrence of misattributed paternity over time, from 3.2% (95% credibility interval, 2.7–3.8%) for children born in the 1930s to 0.9% (95% credibility interval, 0.7–1.1%) for children born in the 2010s (Fig. 1b). The misattributed paternity frequency was highest among those with the lowest educational level for both the mother and father, 2.0% (95% credibility interval, 1.8–2.1%) and 1.7% (95% credibility interval, 1.5–1.8%), respectively, but with a trend towards higher estimates for mothers with postgraduate education 1.1% (95% credibility interval, 0.6–1.8%) (Fig. 1c). For the analysis regarding the age of father and mother, there was a u-shaped association with higher estimated rates for younger and older age (Supporting Information Fig. S2). For geographical regions, the were small or no associations between place of birth and estimated rates for population-dense or non-dense regions with a few exceptions (Supporting Information Fig. S3).

Discussion

In this study, we present population-based estimates of the frequency of misattributed paternity in Sweden from the 1930s until now. Overall, we find misattributed paternity to be within the range

© 2021 The Authors. Journal of Internal Medicine published by John Wiley & Sons Ltd on behalf of Association for Publication of The Journal of Internal Medicine.

Journal of Internal Medicine, 2022, 291; 95–100
The frequency of misattributed paternity in Sweden is low and decreasing / T. Dahlén et al.

Fig. 1 Results from main and subgroup analysis: (a) Main analysis by the Poisson model (black error bars depict 95% confidence intervals) and Bayesian model (black error bars depict 95% credibility intervals). (b) Sub-group analysis by the birth decade of offspring (black error bars depict 95% credibility intervals). (c) Sub-group analysis by the highest educational level of mothers and fathers (black error bars depict 95% credibility intervals).

reported for other countries, with the contemporary rate being approximately 1%, although it may be as high as 3% in persons born in the 1940s, or earlier. In addition to its societal implications, it indicates that misattributed paternity is unlikely to have large effects on studies investigating familial disease risks, especially if these studies are conducted using relational data collected in recent decades.

The study comprises some significant strengths. Most notably, the inclusion of a large portion of the Swedish population, namely those who had undergone blood grouping prior to routine health care procedures, likely renders the results generalizable to the Swedish population. This, taken together with the high quality of the ABO blood group data in the SCANDAT3-S database which is based on largely unprocessed blood group data from transfusion medicine clinics where health consequences of coding errors would be severe, implies that the results should be very robust. Excluding individuals who only had undergone blood grouping testing in preparation for donating blood was done in order to improve generalizability by removing blood donors, a group that is over-represented in the SCANDAT3-S database, who are selected because of good health and a low-risk sexual behaviour and may thus not be representative of the general population. Still, persons undergoing blood group testing in preparation for surgery, childbirth or other medical reasons, may also not be entirely representative, on account of there being a higher disease burden and thus also a higher prevalence of common risk factors for poor health. Conceivably, this may have affected the misattributed paternity frequency if risk factors for poor health are associated with increased or decreased risk of misattributed paternity. Moreover, any specific factor, such as coding errors in the contributing registers, chimerism and other rare reasons for familial blood group mismatch, that could explain the impossible mating patterns in our data would lower the estimates. However, such instances would be very rare. Instead, we believe that our results are limited mainly by the quality of the multi-generation register used to construct the input data for the model. For example, a general improvement in register quality over time might at least partly explain the higher misattributed paternity frequencies in more distant decades. Other reasons for the higher estimates in earlier decades can only be speculated upon, where a possible explanation could be related to an increased tendency for absent fathers due to labour in distant cities or at sea – which was more common previously – or for societal reasons, where attitudes towards divorce were previously less permissive, and where contraceptive methods were less accessible. In general, one could argue that results from analyses of time periods further away may need to be interpreted with greater caution.
The frequency of misattributed paternity in Sweden is low and decreasing / T. Dahlén et al.

The assumption of no gene flow in Hardy-Weinberg equilibrium may be violated due to the extensive recent migration to Sweden slightly altering the expected counts. We have provided the blood group distribution of offspring by the decade of birth in Supporting Information Table S2 to demonstrate small differences in distributions between decades. However, the database contains only a relatively small number of immigrant families limiting the impact of the gene flow effect.

A further issue is a possibility that for some of the offspring we identified as having incompatible blood groups with their parents, it may not be unknown to the recorded parents that the child had a different father and would if so, not be a strict occurrence of misattributed paternity. However, the estimated frequencies are similar to previous estimates and our findings probably reflect an upper bound of misattributed paternity in Sweden.

Hence, we argue that the true, contemporary frequency of misattributed paternity in the Swedish population is close to 1%, although it may be as high as 3% in persons born in the 1940s or earlier. As such, our findings show that misattributed paternity in this large population-based sample is in line with previous contemporary estimates from Europe, further contributing to our understanding of population genetics as well as societal structure.

Funding Information
The creation of the SCANDAT3-S database and the conduct of this study was made possible by a grant to Dr. Edgren from Swedish Research Council (2017-01954). Dr. Edgren is supported by Region Stockholm (clinical research appointment). Dr. Dahlén is supported by Region Stockholm (clinical research appointment). Dr. Zhao is supported by the Clinical Scientist Training Program and the Research Internship Program, both at Karolinska.

Conflict of interest
The authors have declared no conflict of interest.

References
1 Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. Nat Rev Genet. 2008;6:847–59.
2 Bellis MA, Hughes K, Hughes S, Ashton JR. Measuring paternal discrepancy and its public health consequences. J Epidemiol Community Health. 2005;59:749–54.
3 Le Roux MG, Pascal O, Andre MT, Herbert O, David A, Moisan JP. Non-paternity and genetic counselling. Lancet 1992;340:607.
4 Macintyre S, Sooman A. Non-paternity and prenatal genetic screening. Lancet 1991;338:869–71.
5 Sceiza BA, Prall SP, Svinford N, Gopalan S, Atkinson EG, McElreath R, et al. High rate of extrapair paternity in a human population demonstrates diversity in human reproductive strategies. Sci Adv. 2020;6:eaa5195.
6 Yu N, Kruskall MS, Yunis JJ, Knoll JHM, Uhl L, Alosco S, et al. Disputed maternity leading to identification of tetragametic chimerism. N Engl J Med. 2002;346:1545–52.
7 Anderson Kermyt G. How well does paternity confidence match actual paternity? Curr Anthropol. 2006;47:513–20.
8 Book JA. Clinical and genetical studies of hypodontia. I. Premolar aplasia, hyperhidrosis, and canities prematura; a new hereditary syndrome in man. Am J Hum Genet. 1950;2:240–63.
9 Sasse G, Muller H, Chakraborty R, Ott J. Estimating the frequency of nonpaternity in Switzerland. Hum Hered. 1994;44:337–43.
10 Gilding M. Rampant misattributed paternity: the creation of an urban myth. People Place. 2005;13:1–11.
11 Larmuseau MH, Vanoverbeke J, Van Geystelen A, Defraene G, Vanderheyden N, Matthys K, et al. Low historical rates of cuckoldry in a Western European human population traced by Y-chromosome and genealogical data. Proc Biol Sci. 2013;280:20132400.
12 Cokić VP, Kecmanovic M, Zgonjanin Bosić D, Jakovski Z, Veljkovic A, Katic S, et al. A comprehensive mutation study in wide deep-rooted R1b Serbian pedigree: mutation rates and male relative differentiation capacity of 36 Y-STR markers. Forensic Sci Int Genet. 2019;41:137–44.
13 Larmuseau MHD, Claerhout S, Nivelle K, Vandenbosch M, Peeters A, et al. Genetic-genealogy approach reveals low rate of extrapair paternity in historical Dutch populations. Am J Hum Biol. 2017;29:e23046.
14 Wolf M, Musch J, Encezmann J, Fischer J. Estimating the prevalence of nonpaternity in Germany. Hum Nat. 2012;23:208–17.
15 Larmuseau MHD, van den Berg P, Claerhout S, Calafell F, Boattini A, Gruyters L, et al. A historical-genetic reconstruction of human extra-pair paternity. Curr Biol. 2019;29:4102–7.e7.
16 Larmuseau MHD, Matthijs K, Wenseleers T. Cuckolded fathers rare in human populations. Trends Ecol Evol. 2016;31:327–9.
17 Zhao J, Rostgaard K, Hjalgrim H, Edgren G. The Swedish Scandinavian donations and transfusions database (SCANDAT3-S)–50 years of donor and recipient follow-up. Transfusion. 2020;60:3019–27.
18 Ekkom A. The Swedish multi-generation register. Methods Mol Biol. 2011;675:215–20.
19 Adrianto I, Montgomery C. Estimating allele frequencies. Methods Mol Biol. 2012;850:59–76.
Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supporting information ▶