Drug Saf (2013) 36:505–513
DOI 10.1007/s40264-013-0059-1

SHORT COMMUNICATION

Probabilistic Record Linkage for Monitoring the Safety of Artemisinin-Based Combination Therapy in the First Trimester of Pregnancy in Senegal

Stephanie Dellicour · Philippe Brasseur · Per Thorn · Oumar Gaye · Piero Olliaro · Malik Badiane · Andy Stergachis · Feiko O. ter Kuile

Published online: 15 May 2013

Abstract

Background There are insufficient data on the safety in early pregnancy of the artemisinins, a new class of antimalarials. Assessment of drug teratogenicity requires large sample sizes for an adequate risk-benefit assessment. There is currently limited pharmacovigilance infrastructure in malaria-endemic countries. Monitoring drug safety in early pregnancy is especially challenging, as it requires early pregnancy detection to assess any potential increased risk of miscarriage, prospective follow-up to reduce recall and survival biases, and accurate data on gestational age assessment. Record linkage approaches for pregnancy pharmacovigilance using routinely generated health records could be a pragmatic and cost-effective approach for pharmacovigilance in early pregnancy, but has not been evaluated in resource-poor settings.

Objective Our objective was to assess the feasibility of record linkage using routinely collected healthcare data as a pragmatic means of monitoring the safety in early pregnancy of artemisinin-based combination therapies (ACTs) in Senegal.

Methods Data (2004–2008) from paper-based registers from outpatient clinics, antenatal care services (ANC) and the delivery unit from the St Joseph dispensary in Mlomp, south-western Senegal, were entered into databases.

Electronic supplementary material The online version of this article (doi:10.1007/s40264-013-0059-1) contains supplementary material, which is available to authorized users.

S. Dellicour (✉) · F. O. ter Kuile
Child and Reproductive Health Group,
Liverpool School of Tropical Medicine, Pembroke Place,
Liverpool, Merseyside L3 5QA, UK
e-mail: sdell@liv.ac.uk

S. Dellicour
Center for Global Health Research (CGHR),
Kenya Medical Research Institute (KEMRI), Kisumu, Kenya

S. Dellicour
Graduate School, Academic Medical Centre,
Amsterdam, The Netherlands

P. Brasseur
Institut de Recherche pour le Développement (IRD),
Dakar, Senegal

P. Thorn
Thorn IT Services Limited, London, UK

O. Gaye
Faculté de Médecine, Service de Parasitologie,
Université Cheikh Anta Diop, Dakar, Senegal

P. Olliaro
UNICEF/UNDP/WB/WHO Special Programme
for Research and Training in Tropical Diseases (TDR),
Geneva, Switzerland

P. Olliaro
Nuffield Department of Medicine, Centre for Tropical Medicine
and Vaccinology, Churchill Hospital, University of Oxford,
Oxford, UK

M. Badiane
District Médical d'Oussouye, Oussouye, Senegal

A. Stergachis
Departments of Epidemiology and Global Health,
Global Medicines Program, School of Public Health,
University of Washington, Seattle, USA

Record linkage based on a probabilistic matching approach was used to identify pregnancies exposed to ACTs in the first trimester of pregnancy. Two record linkage software packages (Link-Plus and FRIL) were compared and output data were reviewed independently by two investigators.

Results Information on 685 pregnancies was extracted, 536 of which were from the geographic catchment area and eligible for record linkage; 94.6 % of them resulted in live births, 2.6 % in stillbirths and 2.8 % in miscarriages. Major congenital malformations were identified in 1.6 % of births. Seventy-three and 75 true matches between pregnancy outcome and the outpatient treatment registers were identified by two different record linkage software packages, respectively. Record linkage identified seven exposures to ACTs in the first trimester, all of which resulted in normal live-births.

Conclusion Probabilistic record linkage is a potentially cost-effective method to assess the safety of antimalarials in early pregnancy in resource-constrained settings to assess increased risk of overall birth defects, and stillbirths in settings with good existing health records and well defined target populations.

1 Background

In the last decade, Senegal and other malaria-endemic countries changed their first-line treatment policy for malaria to artemisinin-based combination therapy (ACT). The artemisinin class of antimalarials all have embryotoxic effects at low dose ranges in all animal species studied [1, 2] and the World Health Organisation (WHO) does not recommend their use for non-severe malaria in the first trimester as there is insufficient information about their safety in humans. Because of their widespread use, many women in endemic countries risk inadvertent exposure early in pregnancy when they are either unaware of their pregnancy or do not report being pregnant. To date, the data from 359 well documented exposures in the first trimester suggests that the benefits outweigh the potential safety concerns but more data from a wider range of malaria-endemic countries are required to provide adequate reassurance [4–7].

Passive mechanisms of spontaneous adverse drug effects reporting are inadequate to detect drug-induced fetal risks or lack of such risks [3]. Different prospective study designs, including pregnancy registers, are used to monitor safety of medication used during pregnancy in the post-marketing phase [4, 5]; however, these require considerable resources and a well functioning health system and records infrastructure, which are often unavailable in resource-constrained settings [3, 6].

In the last few years there have been an increasing number of pregnancy postmarketing studies using record linkage approaches in developed countries [7–10], which may also have application in resource-limited countries to evaluate the teratogenicity of a drug. It enables rapid evidence generation by using existing healthcare data collected prospectively, while eliminating potential recall bias. Often data on drug exposure, prenatal and pregnancy outcomes are available from multiple linkable data sources. In industrialised countries, this information can be derived from medical records and automated databases, including insurance claims [11]. A recent study showed the feasibility of using record linkage in resource-constrained settings to assess adverse reactions to antiretroviral therapy [12], but such approaches have not yet been applied to assess drug safety in pregnancy in such settings.

Record linkage requires access to long-term, comprehensive and stable population datasets with personal identifiers. In situations where unique identifiers are available, a deterministic record linkage technique, which involves exact matching, can be applied. Probabilistic record linkage is used in situations where there is no universal unique personal identifier and is based on the assessment of similarity between pairs of records, allowing for a level of error (such as typographical or spelling differences) in matching variables [13–15]. The mathematical framework underlying probabilistic record linkage was first formulated by Newcombe and Kennedy in 1962 [16] and further developed by Fellegi and Sunter [13] in 1969. A linkage probability score is computed for each pair of records based on the sum of the probability of agreement for each matching variable. Probability of agreement for each matching variable reflects the probability that the values match by chance based on the frequency of that value occurring in the datasets. Thresholds for the combined linkage probability score are set to determine which pairs are a true match, those that are potential matches and those that are not a match.

We report the results of a study to determine the feasibility of record linkage as a pragmatic approach to retrospectively examine the potential risk associated with inadvertent exposure to ACTs in early pregnancy using data generated through routine clinical practice in a rural mission dispensary in southern Senegal.

2 Methodology

2.1 Study Setting

This study was conducted in 2009 in a mission dispensary based in Mlomp, a rural village of approximately 8,000 inhabitants in the District of Oussouye, Casamance, south-

western Senegal. Malaria is meso-endemic in this area, occurs year round and peaks during the rainy season (July to December). A recent study showed that malaria transmission intensity in southern Senegal has been decreasing significantly in the past 15 years [17].

Mlomp has had a private dispensary operated by French catholic nuns since 1961. The dispensary offers outpatient services, antenatal care service (ANC) once a week, and has a delivery unit. Nearly all pregnant women attend ANC, and health facility deliveries have increased from 50 % in 1961 to 99 % in 1999 [18, 19].

The setting is potentially well suited for record linkage studies. The clinic serves a stable, well defined population and the dispensary registers for ANC, delivery, child welfare clinic and general outpatient visits have been meticulously kept since 1993 [20]. Almost all antimalarials used are provided by the clinic because there is no external source of drugs within the study area; the closest pharmacy or drug store is in the nearest town, Oussouye, 10 km away, with limited options for public transport.

2.2 Data Collection

The general outpatient, antenatal, pregnancy complications/miscarriage and delivery registers from Mlomp dispensary covering the period 2004–2008 were digitalised using a digital camera (Canon EOS 450D with external flash and tripod). The data from the digital images of the registers were subsequently entered into Microsoft® Excel spreadsheets. For the purpose of this feasibility study, only entries for women of childbearing age (15–49 years) with a prescription for an ACT between January 2004 and December 2007 were extracted from the outpatient register. All treatment information, including treatment given at the time of antenatal visits, is recorded in the outpatient register. All recorded pregnancy outcomes between May 2004 and September 2008 were entered from the pregnancy complication and the delivery registers (Table 1). Information on newborn abnormalities was recorded in the delivery register based on observation by the midwife attending the delivery. Data from the outpatient register were double entered and the entries for the delivery register were compared with data extracted for a previous study [20].

2.3 Record Linkage Method

Probabilistic record linkage based on first names, surname, address and year of birth was used to link information from the different registers because they did not contain unique patient identifiers (Table 2). Two stand-alone software packages were used and compared: (i) Registry Plus™ Link Plus (Emory University, v. 2.0) [21] (henceforth referred to as 'Link-Plus'), a royalty-free probabilistic record linkage program developed at the Division of Cancer Prevention and Control of the Centers for Disease Control and Prevention (CDC), USA; and (ii) Fine-grained Record Linkage software (FRIL v. 2.1.5), a free open-source tool developed by Emory University and the CDC [22, 23]. Following agreement on the matching criteria, two investigators conducted the record linkage independently. Any discrepancies in results between investigators were compared and discussed until a consensus was reached.

A variety of matching methods are available for each matching variable to take into account different spelling, recording or typographical errors. To identify the optimal parameters that produce the best linkage results, different linkage setups were compared by varying the weights given to the matching variables. The adequacy of the linkage result was assessed by the distribution of matches, uncertain and non-matches according to the linkage score, the number of true matches identified and the positive predictive value (PPV) for the selected threshold. In Link-Plus, the matching score threshold was set to 0 (minimal threshold option) in order to derive histograms from the linkage output for each matching set up (this was not an option in FRIL). True matches were defined as records with matching first names, surname, and address, allowing for some misspelling and year of birth within 5 years of each other. Pairs were assigned as uncertain matches if first name(s) matched, year of birth was within 10 years of each other but the address and/or surname did not match. Surname and address can change over time, and this is especially common among women of childbearing age when they get married. All other pairs of records were considered as non-match records. Details of the selected linkage setups are provided in Table 3. For information on the linkage optimisation procedures, see appendices I and II in the Online Resource. Deduplication of the delivery register was performed using the deduplication function of Link-Plus, as some women could have multiple pregnancies within the 4-year study period, and to enable one-to-many matching.

2.4 Analysis

Descriptive statistics were used to determine the proportion of women with inadvertent exposure to ACTs during the first trimester of pregnancy using the data obtained by linking the records from the outpatient and the delivery/pregnancy complication registers. Characteristics of women included and excluded from the record linkage were compared using the Pearson Chi-square statistic. A first trimester exposure was defined as an ACT prescription (either artesunate-amodiaquine or artemether-lumefantrine) provided to any woman in 2–14 weeks (inclusive) of

Table 1 Summary of data available in each health register from the Mlomp Dispensary in Senegal

	General outpatient register	Antenatal register	Pregnancy complication register	Delivery register
Identifiers and demographic	• First and last names • Age • Sex • Address	• First and last names • Age • Address	• First and last names • Date of birth/age • Marital status	• First and last names • Date of birth/age • Address • Marital status • Employment
Clinical information and care provided	• Presenting symptoms • Diagnosis • Treatment • Comments	See ANC row 4	• Hospitalisation dates • Care and drugs provided	• Drugs provided during delivery or post-partum to mother and baby
Obstetric history	None	• No. of previous pregnancies • No. of previous deliveries	• No. of pregnancies • No. of live-births • No. of stillbirths • No. of miscarriage • No. of children alive	• Date of delivery • Name of child • Sex • Place of delivery • Child status
ANC	None	• Date of ANC visit • LMP • Weight, Hb, fundal height, blood pressure for each visit • Risk factors • Tetanus vaccination	None	• Date of ANC visits • Weight, Hb, fundal height, blood pressure for each visit • Tetanus vaccination
Delivery	None	None	• Miscarriage information	• Date of delivery • Gestational age at delivery • Baby sex, weight and height • Comments on delivery • Any abnormalities detected for the newborn • Date of discharge

ANC antenatal care, *LMP* last menstrual period

Table 2 Description of matching variables from the delivery register

Matching variables	Discriminating power/# possible values	Missing value (with ANC info)[a]	Variable limitations
First name	230	0	• Variation in spelling/or spelling errors • Truncated/nicknames • Hyphenated/multiple names • Inconsistent and interchangeable use and recording of traditional and Christian first and second names
Surname	73	0.1 %	• Variation in spelling/or spelling errors • Commonly changes for women of childbearing age after marriage/divorce
Address	29	33 % (10 %)	• Changes over time • Commonly changes for women of childbearing age after marriage/divorce • Address only consists of neighbourhood name, which is not very discriminating
Age or year of birth	33	10 % (4 %)	• Inaccuracy in age or date of birth is common in the study population

[a] Manual search of the corresponding records in the ANC register enabled completion of certain missing values for age and address

ANC antenatal care

Table 3 Description of linkage parameters in Link-Plus and FRIL

Linkage variables[a]	Link-Plus set up		FRIL set up[b]	
	M-probability[c]	Matching method[d]	Weight	Matching method[d]
First Name	0.97	First Name (J-W)	32	Edit distance (0.1–0.4)
Second Name	0.7	Generic String	10	Edit distance (0.1–0.4)
Surname	0.85	Last Name (J-W)	25	Edit distance (0.1–0.4)
Address	0.85	Generic String	23	Edit distance (0.1–0.4)
Year of birth	0.2	Generic String	10	Numeric distance (±5 years)

[a] All variables are standardized

[b] FRIL doesn't calculate m-probability as in Link-Plus but enables the user to set weights for each matching variable summing to 100. To compute the total score for a pair, the weight of each matching variable is multiplied by the value returned by chosen edit distance function and all these values are added together

[c] m-probability determines the reliability of the variable, it ranges from 0 (unreliable) to 1 (very reliable)

[d] Matching methods descriptions: J-W: Jaro-Winkler Metric is a string comparator that measures the partial agreement between two strings accounting for random insertion, deletions, and transpositions; Generic String and Edit-distance: incorporates partial matching to account for typographical errors and calculates the number of operations (insertion, deletion, or substitution of a single character) needed to transform one string into the other. The approve and disapprove levels need to be specified in FRIL for edit distance function; numeric distance: allows users to specify a range of values that will have a non-zero match score

pregnancy. Data on last menstrual period (LMP) were not collected in either the outpatient or the delivery registers. Information on the estimated gestational age was only available in the delivery register and based on the assessment by the midwife at birth. This was categorised as 'term' or 'pre-term'. For pre-term deliveries, the gestational age at delivery was specified in months from LMP in the delivery records, whereas for all 'term' deliveries, the specific gestational age was not recorded other than the notation that they were 'term' births. It was therefore set at 40 weeks from LMP for the purpose of the analysis. The gestational age at the time of exposure was derived from the date of the ACT prescription in the outpatient register and the estimated date of conception calculated from the gestational age assessment at birth. Analysis was done in SPSS version 18.

3 Results

3.1 Description

Between May 2004 and September 2008, 685 pregnancy outcomes were captured from the pregnancy complication

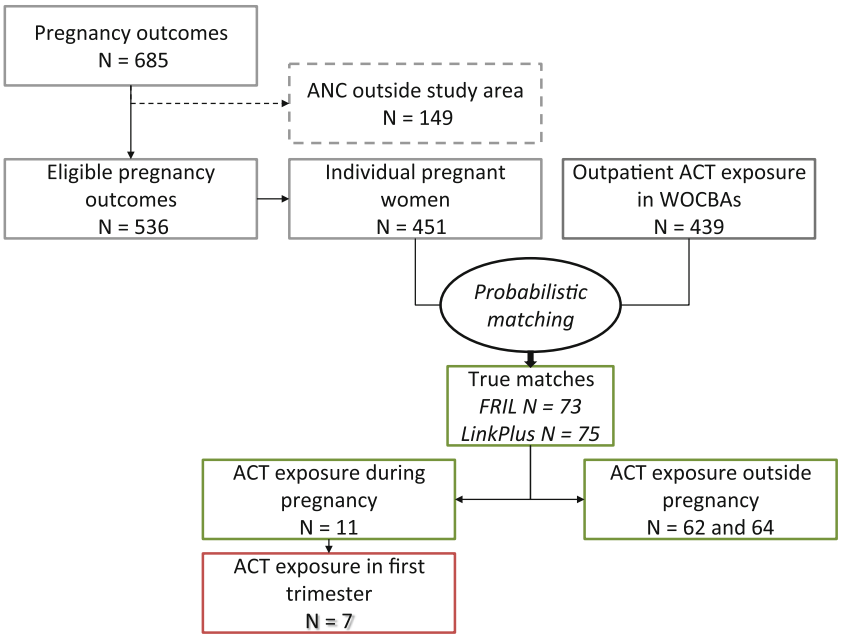

Fig. 1 Flow diagram for inclusion in record linkage, matching result and resulting ACT pregnancy exposure. Note that of the 451 women, 372 had only one pregnancy during the study period (2004–2008), 73 had two pregnancies and six had three pregnancies during the 4-year period

Table 4 Characteristics of 536 eligible pregnancies from the delivery and pregnancy complication registers (2004–2008) St Joseph Dispensary Mlomp, Senegal

Characteristic	N (%)[a]
Age in years	
Mean (SD)	29.2 (6.6)
Range	15–47
Parity	
0	94 (17.5)
1	87 (16.2)
2+	355 (66.2)
Range	0–13
Number reporting previous stillbirth	26 (4.9)
Number reporting previous miscarriage	55 (10.3)
Pregnancy outcomes	
Live births	507 (94.6)
Stillbirths	14 (2.6)
Miscarriages	15 (2.8)
Preterm at delivery	26 (4.9)

[a] Unless otherwise indicated

Table 5 Description of congenital abnormalities identified from the delivery registers (2004–2008) in St Joseph Dispensary Mlomp, Senegal

Congenital abnormalities[a]	Number
Anophthalmia	1
Ambiguous genitalia	1
Anencephaly	3
Down syndrome	6
Club foot	2
Hydrocephalus	1
Imperforate anus	1
Unspecified malformations	1

[a] Some cases had more than one abnormality. There were 16 abnormalities among 11 births: two cases had both Down syndrome and anencephaly, one case of Down syndrome also had a club foot, one case of anophthalmia also had a club foot and one case with Down syndrome also had hydrocephalus

and delivery registers. Women who attended their ANCs outside of the Mlomp catchment area ($n = 149$) were excluded from the record linkage because no outpatient treatment records were available from these other clinics. Their characteristics did not differ significantly (at 5 % significance level) from other women living in the catchment area, except that they were more likely to be primiparous women and to be slightly younger. Figure 1 depicts the number of records that contributed to the probabilistic matching.

Of the pregnancies included, 94.6 % resulted in live births, 2.6 % in stillbirths and 2.8 % in miscarriages (Table 4). Overall, 11 cases of birth defects (1.6 %) were captured in the delivery register (Table 5).

3.2 Probabilistic Matching

The optimum set up with highest PPV (86 %) and highest number of matches was accomplished by including second names. Figure 2 depicts the distribution of match, uncertain and non-match for this set up. The relatively high number of false negatives ($n = 13$) and false positives ($n = 8$) suggests clerical review would still be required for pairs with a linkage probability score between 6 and 9.

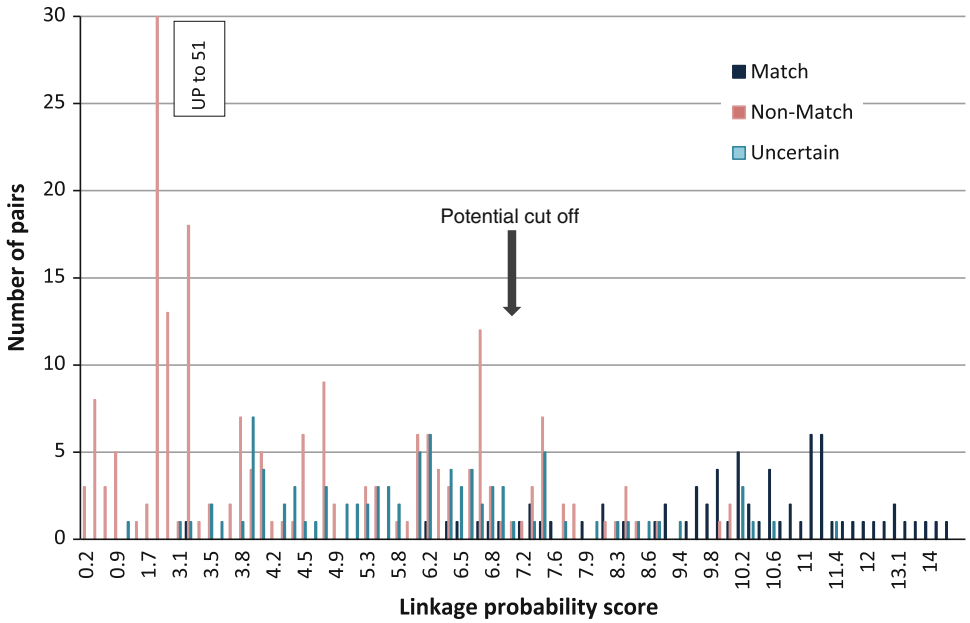

Fig. 2 Histogram of the optimum set up depicting the distribution of matches, uncertain and non-matches according to the probability score derived from Link-Plus

Seventy-one and 75 matched pairs between the pregnancy outcome registers and the ACT treatment data from the outpatient registers were detected using Link-Plus and FRIL, respectively. The four additional matches identified through FRIL were not detected through Link-Plus as it only allows one-to-many matching. After running another round of matching in Link-Plus using the non-match records only and adjusting the matching variables weight, an additional two matches were identified (total of 73 matched pairs). The other two treatment records were matched by Link-Plus with incorrect records from the delivery register and were not compared again to the correct delivery records. The two cases missed in Link-Plus were not exposed during the pregnancy period.

3.3 Artemisinin Exposure in Pregnancy

Out of the true matches, 11 of the 536 pregnancies (2 %) had evidence for ACT exposure during pregnancy, seven during the first trimester (1 %), three of whom were 4–10 weeks (the projected embryo-sensitive period from animal models) pregnant (0.6 %) and four in the second or third trimester. There was no exposure in the month prior to the estimated conception dates. None of the women exposed during pregnancy had an adverse pregnancy outcome, preterm birth or a baby with malformations.

4 Discussion

The findings from this feasibility study suggest that record linkage using routine healthcare data is feasible in resource-constrained settings with a relatively well defined catchment population. Among the 451 women with pregnancy outcomes, 75 records from the delivery register were matched to an ACT treatment in the outpatient registers, although only 11 were ACT exposures during the pregnancy period; the other matches represent ACT treatment received before or after pregnancy. Matching was feasible despite the limited discriminating power resulting from the lack of variance among two of the five matching variables due to clustering of commonly used surnames, lack of address details and the lack of precise dates of birth for many individuals. Furthermore, the interchangeable use of traditional and Christian first names by single individuals was very common and often one was recorded, but seldom both.

Although there was no other source of medication in Mlomp and the nearby surrounding areas, the risk of exposure misclassification remains, as women could have obtained drugs from relatives or friends, could have travelled during their pregnancy and obtained drugs elsewhere or the drugs could have been prescribed and recorded in the register but never actually taken by the patient. The limited detail on gestational age (categorised as term or pre-term with gestational month) could also cause misclassification for timing of exposure. Furthermore, there could have been errors or omissions during recording of data in the health registers and at time of data entry. We attempted to minimise the latter by using double data entry for the general outpatient register and by comparing the entries for the delivery register with data extracted for a previous study [20].

The prevalence of birth defects was within the expected range for major malformations detectable at birth by surface exams by non-specialists without special training for newborn examination. Although the background prevalence of

birth defects in developing countries is unknown, extrapolation from birth defect prevalence found in industrialised countries suggests that the defect prevalence detectable at birth would be around 1 % after exclusion of defects of genetic aetiology and heart defects, which require specialist assessment to be diagnosed. The background rate of stillbirth (2.3 %) was also within the projected range for Senegal (2.7 %) and sub-Saharan Africa (3.2 %) [24]. As expected, the proportion of miscarriages (2.5 %) was much lower than the risk of ~12–15 % typically quoted for miscarriages because most pregnant women do not present for their first ANC visit before the second trimester. A previous record linkage study from Mlomp dispensary showed that half of all pregnant women presented for their first ANC visit after 20 weeks gestation, thus only allowing the prospective recording of miscarriages relatively late in the second trimester (20–28 weeks) [20].

The two record linkage software programs used had comparable performance. The main advantage of FRIL was that it included a matching method for numeric variables, which allowed variation in numeric values (i.e. year of birth could be within 5 years of each other and considered a match). There is no function for numeric variables in Link-Plus, and year of birth had to be treated as a string variable. Many-to-many pairwise comparison is possible in FRIL, whereas only one-to-many is enabled in Link-Plus (many-to-many will be available in the next version, 3.0 [25]). This was highlighted by the two missed true matches identified through FRIL but not Link-Plus described above. Limitations for FRIL included the more limited flexibility to export datasets after linkage (only matched pairs can be exported). Link-Plus, on the other hand, enables the export of all pairs reviewed manually (matches, uncertain and non-matches), which then allows more complete secondary analysis and graphical depiction of results.

Ruling out the teratogenic risk of a drug requires a very large sample size, as teratogens usually induce specific patterns of birth defects that occur rarely in the general population. We have shown previously that 10,748 well characterised exposures and four times as many unexposed controls would be required to exclude a doubling of risk of a specific birth defect that occurs at a frequency of 0.1 % [11]. Obtaining reliable data on early pregnancy drug exposures is challenging. McGready and colleagues recently reported 64 well documented first trimester exposures to ACTs after reviewing 25 years of data on 48,426 pregnancies [26]. This was enough to exclude a doubling of risk of miscarriage for first trimester exposures [27]. To obtain a similar level of reassurance for major malformations, data from several sentinel sites over several years are required. Collaborating pregnancy registry sites have been set up by WHO [28] and the Malaria in Pregnancy consortium [29] for this purpose.

Record linkage studies require minimal staff (for data extraction, data entry and analysis) and resources (i.e. digital camera and a few computers for data entry). In settings with electronic medical records, only staff for data management and analysis would be required. However, assessment of the risk of miscarriage and specific birth defects, such as congenital heart defects, requires dedicated studies. A 'miscarriage' endpoint requires an observational cohort study involving women of childbearing age to capture data as early as possible in their pregnancy, while assessment of specific rare birth defects would probably require nested case control or case cohort studies in settings where drug exposure information is captured routinely. Additional data from sites where record linkage is a possibility would greatly contribute to the achievement of an adequate sample size to guide policy makers. Requirements for such sites include availability of reliable and comprehensive medical records for treatment, pregnancy and maternity services, in a relatively stable population where healthcare is provided in a central location with a high level of health facility deliveries and limited availability of the drug of interest outside the central health facility. Settings such as private agricultural estates (e.g. tea, coffee, sugar or cotton plantations), where employees and their families get healthcare centrally and routine healthcare data are typically recorded, could be used for similar record linkage studies. Data from health insurance schemes, where unique identifiers are available, could be used with a combination of deterministic and probabilistic record linkage [30, 31].

5 Conclusion

Our findings suggest that record linkage to assess drug safety in pregnancy in resource-constrained settings is feasible for assessment of stillbirths and major congenital malformations detectable by surface examination. Tapping into readily available data sources of sites adequate for record linkage would greatly contribute to the high numbers needed to provide adequate reassurance for ACT use in the first trimester of pregnancy.

Acknowledgments This work was made possible by the dedication of the healthcare personnel of St Joseph Dispensary in Mlomp and the help of Sister Marie Joelle.

Funding SD is partly funded by the Malaria in Pregnancy Consortium, which is funded through a grant from the Bill & Melinda Gates Foundation to the Liverpool School of Tropical Medicine. SD is also grateful to the US Centers for Disease Control and Prevention for salary support through a cooperative agreement between the Division of Parasitic Diseases and Malaria (Centers for Disease Control and Prevention, USA) and the Malaria Epidemiology Unit of the Child and Reproductive Health group, Liverpool School of

Tropical Medicine held by FtK. Travel expenses for SD and PT were covered by a pharmacovigilance planning grant from the Bill & Melinda Gates Foundation to the University of Washington held by AS.

Conflict of interest The authors declare no conflict of interest. PO is a staff member of the WHO; the authors alone are responsible for the views expressed in this publication and they do not necessarily represent the decisions, policy or views of the WHO.

Author's contributions SD, FtK, OG and PO contributed to the concept of the project. SD and FtK developed the protocol with contributions from PO, PB and AS. SD and PT conducted the data collection and record linkage analyses. PB supervised the data collection. SD and FtK wrote the first draft of the manuscript; all authors reviewed and revised the final version.

References

1. WHO. Assessment of the safety of artemisinin compounds in pregnancy. Report of two informal consultations convened by WHO in 2002. Geneva: WHO; 2003.
2. WHO, TDR. Assessment of the safety of artemisinin compounds in pregnancy: report of two joint informal consultations convened in 2006 by The Special Programme for Research and Training in Tropical Diseases (TDR) sponsored by UNICEF/UNDP/World Bank/WHO. Geneva: WHO; 2006.
3. Kennedy DL, Uhl K, Kweder SL. Pregnancy exposure registries. Drug Saf. 2004;27(4):215–28.
4. Källén BA. Methodological issues in the epidemiological study of the teratogenicity of drugs. Congenit Anom (Kyoto). 2005; 45(2):44–51.
5. Hernandez-Diaz S, Hernan MA, Meyer K, Werler MM, Mitchell AA. Case-crossover and case-time-control designs in birth defects epidemiology. Am J Epidemiol. 2003;158(4):385–91.
6. Adam MP, Polifka JE, Friedman JM. Evolving knowledge of the teratogenicity of medications in human pregnancy. Am J Med Genet C Semin Med Genet. 2011;157(3):175–82.
7. Andrews EB, Tennis P. Promise and pitfalls of administrative data in evaluating pregnancy outcomes. Pharmacoepidemiol Drug Saf. 2007;16(11):1181–3.
8. Charlton RA, Weil JG, Cunnington MC, Ray S, de Vries CS. Comparing the General Practice Research Database and the UK Epilepsy and Pregnancy Register as tools for postmarketing teratogen surveillance: anticonvulsants and the risk of major congenital malformations. Drug Saf. 2011;34(2):157–71.
9. Colvin L, Slack-Smith L, Stanley FJ, Bower C. Pharmacovigilance in pregnancy using population-based linked datasets. Pharmacoepidemiol Drug Saf. 2009;18(3):211–25.
10. Irvine L, Flynn RW, Libby G, Crombie IK, Evans JM. Drugs dispensed in primary care during pregnancy: a record-linkage analysis in Tayside, Scotland. Drug Saf. 2010;33(7):593–604.
11. Dellicour S, Ter Kuile FO, Stergachis A. Pregnancy exposure registries for assessing antimalarial drug safety in pregnancy in malaria-endemic countries. PLoS Med. 2008;5(9):e187.
12. Corbell C, Katjitae I, Mengistu A, Kalemeera F, Sagwa E, Mabirizi D, et al. Records linkage of electronic databases for the assessment of adverse effects of antiretroviral therapy in sub-Saharan Africa. Pharmacoepidemiol Drug Saf. 2012;21(4):407–14.
13. Fellegi IP, Sunter AB. A theory for record linkage. J Am Stat Assoc. 1969;64(328):1183–210.
14. Jaro MA. Probabilistic linkage of large public health data files. Stat Med. 1995;14(5–7):491–8.
15. Meray N, Reitsma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. J Clin Epidemiol. 2007;60(9):883–91.
16. Newcombe H, Kennedy J. Record linkage: making maximum use of the discriminating power of identifying information. Commun ACM. 1962;5(11):563–6.
17. Brasseur P, Badiane M, Cisse M, Agnamey P, Vaillant MT, Olliaro PL. Changing patterns of malaria during 1996–2010 in an area of moderate transmission in southern Senegal. Malar J. 2011; 10:203.
18. Enel C, Pison G, Lefebvre M. De l'accouchement traditionnel à l'accouchement moderne au Sénégal. Cahiers Santé. 1993;3: 441–6.
19. Pison G, Kodio B, Guyavarch E, Etard J-F. La mortalité maternelle en milieu rural au Sénégal. Population. 2000;55(6): 1003–18.
20. Olliaro PL, Delenne H, Cisse M, Badiane M, Olliaro A, Vaillant M, et al. Implementation of intermittent preventive treatment in pregnancy with sulphadoxine/pyrimethamine (IPTp-SP) at a district health centre in rural Senegal. Malar J. 2008;7:234.
21. US Centers for Disease Control and Prevention. Registry Plus™ Link Plus technical information and installation. Atlanta (GA): CDC; 2006 [online]. http://www.cdc.gov/cancer/npcr/tools/registryplus/lp_tech_info.htm. Accessed 24 Apr 2013.
22. Jurczyk P. Fine-grained record integration and linkage tool. Open source record linkage software. 2.1.4 edn. Atlanta (GA), 2009.
23. Jurczyk P, Lu JJ, Xiong L, Cragan JD, Correa A. Fine-grained record integration and linkage tool. Birth Defects Res A Clin Mol Teratol. 2008;82(11):822–9.
24. Stanton C, Lawn JE, Rahman H, Wilczynska-Ketende K, Hill K. Stillbirth rates: delivering estimates in 190 countries. Lancet. 2006;367(9521):1487–94.
25. NPCR. Link Plus Development Priority List (updated January 26, 2011). 2011. http://www.cdc.gov/cancer/npcr/tools/registryplus/lp_features.htm. Accessed 7 Feb 2012.
26. McGready R, Lee SJ, Wiladphaingern J, Ashley EA, Rijken MJ, Boel M, et al. Adverse effects of falciparum and vivax malaria and the safety of antimalarial treatment in early pregnancy: a population based study. Lancet Infect Dis. 2012;12(5):388–96.
27. Desai M, Dellicour S. Effects of malaria and its treatment in early pregnancy. Lancet Infect Dis. 2012;12(5):359–60.
28. Gomes M. New global pregnancy register established. 2009 [online]. http://www.who.int/maternal_child_adolescent/news_events/news/2009/pregnancy_register/en/index.html. Accessed 15 May 2012.
29. MiPc. Malaria in Pregnancy Consortium. 2012. http://www.mip-consortium.org/. Accessed 2012.
30. Romitti PA, Watanabe-Galloway S, Budelier WT, Lynch CF, Puzhankara S, Wong-Gibbons D, et al. Identification of Iowa live births in the Agricultural Health Study. Arch Environ Occup Health. 2010;65(3):154–62.
31. Roos LL Jr, Wajda A, Nicol JP. The art and science of record linkage: methods that work with few identifiers. Comput Biol Med. 1986;16(1):45–57.