PLOS ONE

RESEARCH ARTICLE

Improved privacy preserving method for periodical SRS publishing

Wei Huang[1], Tong Yi[2]*, Haibin Zhu[3], Wenqian Shang[4], Weiguo Lin[4]

1 Division of Scientific Research, Communication University of China, Beijing, China, **2** School of computer information and Engineering, Guangxi Normal University, Guilin, Guangxi, China, **3** Department of Computer Science and Mathematics, Nipissing University, North Bay, Ontario, Canada, **4** School of Computer Science, Communication University of China, Beijing, China

* yitong@mailbox.gxnu.edu.cn

Citation: Huang W, Yi T, Zhu H, Shang W, Lin W (2021) Improved privacy preserving method for periodical SRS publishing. PLoS ONE 16(4): e0250457. https://doi.org/10.1371/journal.pone.0250457

Editor: M. Usman Ashraf, University of Management and Technology, PAKISTAN

Received: November 8, 2020

Accepted: April 6, 2021

Published: April 22, 2021

Data Availability Statement: The experiment data files are available from the FAERS dataset: https://open.fda.gov/data/faers/ The experimental dataset can be also found at https://doi.org/10.6084/m9.figshare.14269244 The other relevant data are within the paper.

Funding: This paper is partly supported by "Fundamental Research Funds for the Central Universities" , Doctoral research start-up fund of Guangxi Normal University(RZ1900006676)" and "The project of improving the basic scientific research ability of young and middle-aged teachers

Abstract

Spontaneous reporting systems (SRSs) are used to collect adverse drug events (ADEs) for their evaluation and analysis. Periodical SRS data publication gives rise to a problem where sensitive, private data can be discovered through various attacks. The existing SRS data publishing methods are vulnerable to Medicine Discontinuation Attack(*MD*-attack) and Substantial symptoms-attack(*SS*-attack). To remedy this problem, an improved periodical SRS data publishing—PPMS(k, θ, α)-bounding is proposed. This new method can recognize *MD*-attack by ensuring that each equivalence group contains at least k new medicine discontinuation records. The *SS*-attack can be thwarted using a heuristic algorithm. Theoretical analysis indicates that PPMS(k, θ, α)-bounding can thwart the above-mentioned attacks. The experimental results also demonstrate that PPMS(k, θ, α)-bounding can provide much better protection for privacy than the existing method and the new method dose not increase the information loss. PPMS(k, θ, α)-bounding can improve the privacy, guaranteeing the information usability of the released tables.

1. Introduction

Many developed countries have established spontaneous reporting systems (SRSs) for the collection of adverse drug events (ADEs). These datasets allow researchers to analyze possible correlations between drugs and adverse reactions. Typical spontaneous reporting systems include FAERS of the US Food and Drug Administration [1] and the UK Yellow Card scheme [2].

However, these datasets usually involve information which relates to an individual's privacy. Sensitive attributes (SAs), e.g., adverse drug reaction and disease type are also included. Publishers usually remove attributes which can identify individuals uniquely before releasing their reports, however, Sweeney [3] has pointed out that an adversary can use quasi-identification attributes (QIAs) to link the released table to other publicly available datasets in an effort to uniquely identify an individual. A QIA can be Age, Gender, etc. A single quasi-identification attribute cannot uniquely identify an individual. To protect SAs, privacy preserving data publishing (PPDP) usually anonymizes original tables before releasing. In recent years, PPDP has been widely studied and seeks to maintain the tradeoff between privacy/security and

in Guangxi Universities(2020KY020323)". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist

information usability in released tables. *K*-anonymity [3] and its variants [4–8] are only suitable for static tables. When faced with Dynamic data tables, some incremental data publishing methods [9–14] are presented, such as *BCF*-anonymity [9], *m*-invariance [10], etc. Most techniques cannot preserve identity in released tables and are not well suited to SRS data publishing. Differential privacy [15–18] can make presence or absence of a record in the dataset have little effect on the outcome. However, the utility of released tables will be adversely affected by the added noise [19].

SRSs release updated datasets periodically, for example, the US Food and Drug Administration releases the adverse drug event datasets quarterly. Lin et al. [20–21] showed that SRS datasets usually include some characteristics, e.g., multiple individual records, multivalued sensitive attributes, etc. More importantly, the related ADE records of an individual may be contained in tables released in each period. These records share case identification (CaseID) to trace follow-ups to an event [22]. Thus, conventional data publishing methods cannot handle SRS datasets. To resolve this, Wang et al. [22] defined three types of attacks in SRS dataset publishing and presented a periodical SRS data publishing method- PPMS(k, θ^*)-bounding. These attack types are defined as follows.

Definition 1 (Backward-attack)

Assume target individual P whose record t is in sanitized released table T_i, use $t.QIA$ and C to represent the QIAs values of t and P's candidate CaseID set in T_i, respectively. U contains all this record r: r is from the previous released tables $\{T_1, T_2, \ldots, T_{i-1}\}$, and r' s CaseID is in C. The Backward-attack (B-attack) may happen if there is a record r ($r \in U$) whose QIAs values $r.QIA$ does not cover $t.QIA$. We Denote the set of these excludable records as B. The QIAs values of r ($r.QIA$) cover the ones of t ($t.QIA$), if the records r and t satisfy: for each quasi-identification attribute QI in QIA, $r.QI$ is equal to or more generalized than $t.QI$.

Definition 2 (Forward-attack)

Assume target individual P whose record t is in sanitized released table T_i. We use $t.QIA$ and C to represent the QIAs values of t and P's candidate CaseID set in T_i, respectively. U contains all this record r: r is from the subsequent released tables $\{T_{i+1}, T_{i+2}, \ldots\}$, and r's CaseID is in C. The Forward-attack(F-attack) may happen if there is a record r ($r \in U$) whose QIAs value $r.QIA$ does not cover $t.QIA$. Denote the set of these excludable records as F.

Definition 3 (Latest-attack)

Assume target individual P whose record t is in sanitized released table T_i, and none of the previous released tables $\{T_1, T_2, \ldots, T_{i-1}\}$ contain the CaseID of P. We use C to represent the P's candidate CaseID set in T_i. The Latest-attack(L-attack) may happen if there is any CaseID cid ($cid \in C$) which appears in some previous released tables. Denote the set of these excludable records as L.

Based on the above three attacks, the definition of an anonymity model can be introduced [22]:

Definition 4 (PPMS(k, θ^*)-bounding)

Assume $\{s_1, s_2, \ldots, s_m\}$ are all the possible sensitive attribute values in datasets, accordingly, $\theta^* = \{\theta_1, \theta_2, \ldots, \theta_m\}$ are the probability thresholds set by the data holder. Released tables $T_1, T_2, \ldots, T_j$ satisfy PPMS(k, θ^*)-bounding if each T_i ($1 \leq i \leq j$)satisfies:

1. For each individual P, assume the candidate CaseID set of P in T_i is C, then there is $|C-(B \cup F \cup L)| \geq k$;
2. For every individual P, an adversary can conclude that P has any sensitive attribute value s_j with a probability at most θ_j.

PPMS(k, θ^*)-bounding uses *NC*-bounding and *OID*-bounding to defend against the above three attacks. *NC*-bounding makes each group contain at least k new CaseIDs, which can defend against backward-attack and latest-attack. Let t be a record in released T_i, t_1, t_2, . . ., t_j are records have the same CaseID with t in previous tables T_x, T_y, . . ., $T_z(x<y<\ldots,<z)$. To thwart forward-attack, *OID*-bounding requires that QIAs values of t should cover all records that share the same CaseID with t in previous tables. Wang et al. [22] also found that the forward-attack can be avoided when QIAs values of t only cover the records that share the same CaseID with t in T_x, they call this method PPMS_EAR. Their experiments show that PPMS_EAR can thwart the above three attacks while maintaining the usability of tables.

Table 1(A)–1(C) are three original tables, Adverse Drug Reaction (ADR) is a sensitive attribute, and Sex and Age are quasi-identification attributes. For simplicity, letters are used to denote a type of ADR. Table 2(A)–2(C) are corresponding sanitized released tables which satisfy PPMS(3, 1/3)-bounding. Obviously, the released tables can thwart backward-attack, forward-attack and latest-attack. However, an adversary can still disclose the privacy of individuals. Let us consider some examples.

Example 1

Bob knows that Alice(F, 39) is in Table 2(B), and he can relate Alice to records$\{t_{13}, t_{15}, t_{18}\}$. He also knows that Alice will stop medicine in quarter 3, because her illness is shown as cured in quarter 2. Therefore, Bob can exclude t_{13}, t_{15}, and conclude Alice's record is t_{18} with the probability of 100%. The privacy of Alice is disclosed.

Example 2

Bob knows that Clare(M, 47) is in Table 2(B), and can be related to records$\{t_{16}, t_{17}, t_{21}\}$. Bob cannot relate Clare to a unique record, but $\{t_{16}, t_{17}, t_{21}\}$ all contain many more adverse drug reactions than other records. Thus, Bob can conclude that Clare gets more adverse drug reactions than other people. The privacy of Clare is disclosed.

PPMS(k, θ^*)-bounding has not considered medicine discontinuation and records with massive symptoms, hence privacy may be disclosed by an adversary. As the extended versions of PPMS(k, θ^*)-bounding, the other existing SRS data publishing methods [23–25] have not considered the attacks described by example 1 and example 2, either. It is necessary to find a way to defend against the above attacks. However, increasing the security usually makes the information usability decline. It is challenging to balance privacy security against the information utility. To alleviate these problems, this paper proposes a new SRS data publishing method which can improve the privacy and guarantee the information usability. The main contributions of this paper are summarized as follows:

1. Identifying two new attacks which are aimed at ADEs data publishing;
2. Based on the PPMS(k, θ^*)-bounding, proposing a new data publishing method- PPMS(k, θ, α)-bounding. The new method can enhance the security of privacy and preserve the quality of released tables. A corresponding algorithm is presented.
3. Using a real FAERS database from the US Food and Drug Administration to verify PPMS (k, θ, α)-bounding.

Table 1. Three consecutive quarters of original SRS dataset.

(a) Quarter 1

CaseID	Sex	Age	ADR
1	M	50	c,b
7	M	48	a
3	M	46	d
5	M	46	e,g
2	F	21	c,a
4	F	23	b,d
6	F	25	y

(b) Quarter 2

CaseID	Sex	Age	ADR
1	M	50	c,b
11	M	50	a
12	M	53	y
14	M	48	h
16	M	48	q,d,e,p,x
15	F	40	x
18	F	39	q
17	M	46	y,b,c,f,g,i
19	F	43	o
3	M	46	d
20	F	40	x,i
21	F	46	j,z,v,u,k,n
22	M	46	q,l
13	F	39	h,o

(c) Quarter 3

CaseID	Sex	Age	ADR
13	F	40	h,k
26	F	45	x,u
28	F	45	d,i
23	F	40	z
15	F	39	x
27	M	38	a,c
24	M	38	d
25	M	38	e,q

https://doi.org/10.1371/journal.pone.0250457.t001

2. Related work

ADE reporting is a special style of incremental data publishing released periodically. Subsequent tables may add new records, and delete/update records offered previously. For the purpose of tracing individuals, the same CaseID can appear in different released tables. Wang et al. [22] divided traditional incremental data publishing into two types: continuous data publishing and dynamic data publishing.

Continuous data publishing [9, 11]: periodic publishing that carries over records from previously released tables. The data holder needs to release all the data collected so far, if he wants to publish the data which is collected recently. Suppose that the data holder has collected data D_i in timestamped t_i. In general, the data holder has to release R_i which is the anonymized version of $D_1 \cup D_2 \cup \ldots D_i$ in timestamped t_i.

Table 2. PPMS(3,1/3)-bounding publishes tables for Table 1.

(a) Quarter 1				
CaseID	Sex	Age	ADR	Group
1	M	[46–50]	c,b	1
7	M	[46–50]	a	1
3	M	[46–50]	d	1
5	M	[46–50]	e,g	1
2	F	[21–25]	c,a	2
4	F	[21–25]	b,d	2
6	F	[21–25]	y	2
(b) Quarter 2				
CaseID	Sex	Age	ADR	Group
1	M	[48–53]	c,b	1
11	M	[48–53]	a	1
12	M	[48–53]	y	1
14	M	[48–53]	h	1
13	F	[39–40]	h,o	2
15	F	[39–40]	x	2
18	F	[39–40]	q	2
16	*	[46–48]	q,d,e,p,x	3
17	*	[46–48]	y,b,c,f,g,i	3
21	*	[46–48]	j,z,v,u,k,n	3
3	*	[40–46]	d	4
20	*	[40–46]	x,i	4
19	*	[40–46]	o	4
22	*	[40–46]	q,l	4
(c) Quarter 3				
CaseID	Sex	Age	ADR	Group
13	F	[39–45]	h,k	1
26	F	[39–45]	x,u	1
28	F	[39–45]	d,i	1
23	F	[39–45]	z	1
15	*	[38–40]	x	2
27	*	[38–40]	a,c	2
24	*	[38–40]	d	2
25	*	[38–40]	e,q	2

https://doi.org/10.1371/journal.pone.0250457.t002

Some matching records can be excluded by the adversary, because he/she can infer that the records are not related to the target's QIA values or timestamp [9]. Thus, Fung et al. [9] pointed out that the excluded records can help the adversary access to a smaller set of candidates. Thus, they presented a privacy model (called *BCF*-anonymity) to evaluate anonymity after excluding some matching records. Besides, an efficient algorithm was presented to achieve a suboptimal *BCF*-anonymization.

Pei et al. [11] pointed out that in continuous data publishing scenario, *k*-anonymity [3] may be compromised due to the possible inferences using multiple releases. They presented a privacy preserving approach, called Monotonic Incremental Anonymization, to guarantee the *k*-anonymity on each release. Meanwhile, the approach can reduce information loss by using more and more accumulated data.

Some continuous data publishing methods can preserve the identities of individuals among different tables [11], but this type of methods cannot support the operations of deletion and updating. Therefore, dynamic data publishing methods were presented later.

Dynamic data publishing [10, 12–14]: periodic publishing where records can be added, deleted or updated from previously released tables. This method cannot preserve the identities of individuals among different tables. Suppose that the data holder had collected the initial set of tuples D_1 in time t_1, and published R_1 as the anonymized version of D_1. During the period $[t_1,t_2]$, when there were new records coming, the data holder inserted them into D_1. At the same time, some records from D_1 might be deleted or updated by the data holder. Finally, the D_2 could be obtained in time t_2. Thus, the data holder published R_2 as the anonymized version of D_2. In general, the data holder publishes R_i as the anonymized version of D_i in time t_i.

Xiao et al. [10] found out that when incremental data publishing supported deletions, the adversary could disclose the privacy of victims by comparing the series of released k-anonymous [3] and l-diverse [8] data. They presented a privacy model, called m-invariance, to guarantee certain "invariance" in all the QIA groups that a tuple is incorporated into at different publication timestamps.

Li and Zhou [12] defined the updates on attribute values as internal updates. They pointed out that the internal updates related to sensitive values were not arbitrary, the requirement of m-invariance was unreachable in this scenario. A counterfeit generalization approach called m-Distinct was presented to guarantee the security of dynamic publication with internal updates, insertions and deletions.

Following the work of [12], Anjum and Raschia [13] further assumed that new values might not have any association with the old ones, and the adversary knew the "event list". An attack model based on their assumption, called τ- attacks, was defined. To prevent the new attack, Anjum and Raschia also presented a publication approach called τ-safety, which is based on m-invariance and individual-oriented protection.

Bewong et al. [14] illustrated that the transactional data had some special features, such as having many common private terms. Thus, they pointed out that the existing incremental publishing methods were inapplicable. A transactional data publication mechanism called Sanony was also presented, to prevent composition attacks by utilizing counterfeits.

Fully data evolution is supported in dynamic data publishing. However, identity preservation cannot be supported in this type of multiple releases, which results in its inapplicability to ADEs data publishing.

Differential privacy [15–18] has garnered a lot of attention in recent years, it can minimize the chances of identifying records. However, the noise added by differential-based methods is unbounded and random, which will adversely affect the utility of released tables [19].

2016, Lin et al. [21] began to study ADEs data publishing, and presented the MS(k, θ^*)-bounding method based on the characteristics of ADE data. Because this method had not considered the correlation among different released tables, an adversary can exploit this situation when seeking to disclose the privacy of individuals. To resolve this, 2017 Wang et al. [22] proposed PPMS(k, θ^*)-bounding for periodical ADEs data publishing. This method can defend against the three attack methods(B-attack, F-attack and L-attack) which are based on correlations among different released tables. After that, several ADEs data publishing methods based on PPMS(k, θ^*)-bounding were presented [23–25]. Hsiao et al. [23] presented a privacy model, called Closed l-diversity, to process the missing value by guaranteeing that each partial QID-group includes at least l different sensitive values. They also proposed an algorithm, called Closed l-diversification, to achieve Closed l-diversity. Cui et al. [25] presented a SRSs data publication approach, called EQZS, to improve the efficiency of PPMS(k, θ^*)-bounding.

The new values and old values covered each other in this method, which resulted in the limitation of the released data usability.

The existing SRSs data publishing methods have not considered the situation of medicine discontinuation and massive symptoms. The adversary can use related background knowledge to disclose privacy. This paper presents a new ADEs data publishing method to address the matter.

3. PPMS(k, θ, α)-bounding model

In PPMS(k, θ^*)-bounding [22], the adversary learns target individual P's QIAs values, and knows P in a released table. An initial adverse drug reaction can also be revealed. We assume the adversary may learn extra information: P stops medication in the next quarter. Patients will stop the medication when the illness is cured or other therapies(e.g., surgery, food therapy) are chosen. Thus, the assumption is realistic. The adversary can use the information of medication discontinuation to disclose the privacy, like in example 1.

Definition 5 (Medication Discontinuation-attack)

Assume target individual P whose record t is in sanitized released table T_i, and T_{i+1} does not contain the CaseID of P. Use C to represent the P's candidate CaseID set in T_i. The Medication discontinuation-attack(*MD*-attack) may happen if there is any CaseID cid($cid \in C$), which appears in T_{i+1}. Denote the set of these excludable records as MD. Example 1 is an instance of *MD*-attack.

Definition 6 (Substantial Symptoms-attack)

The adversary can conclude that the target individual experiences more symptoms/adverse drug reactions than other people. We have used example 2 to illustrate this style of attack (*SS*-attack). We refer to the record with substantial symptoms/adverse drug reactions as an **ss-record**. Publishers can decide the specific method for defining an ss-record.

Based on PPMS(k, θ^*)-bounding, PPMS(k, θ, α)-bounding needs to thwart these two new attacks:

Definition 7 (PPMS(k, θ, α)-bounding)

Assume $\{s_1, s_2, \ldots, s_m\}$ are values of all the possible sensitive attributes in the datasets, accordingly, $\theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$ are the probability thresholds set by the data holder. Released tables $T_1, T_2, \ldots, T_j$ satisfy PPMS(k, θ, α)-bounding if each T_i ($1 \leq i \leq j$)satisfies:

1. k-bounding: For each individual P, assume that the candidate CaseID set of P in T_i is C, then there is $|C\text{-}(B \cup F \cup L \cup MD)| \geq k$;
2. θ-bounding: For every individual P, an adversary can conclude that P has any sensitive attribute value s_j with a probability at most θ_j.
3. a-bounding: For every individual P, an adversary can conclude that P has many more symptoms/adverse drug reactions than others with the probability at most α.

The privacy requirement of Definition 7(1) is used to avoid record disclosure. *MD*-attack is considered under this requirement which is the extended version of PPMS(k, θ^*)-bounding. The adversary cannot distinguish the target individual from at least k records, even though he/she has excluded some candidates through *MD*-attack, *F*-attack, *B*-attack and *L*-attack. The privacy requirement of Definition 7(2) states that the probability of attribute disclosure will

not exceed a threshold, even though the adversary can exclude some candidate records from various attacks. Thus, this privacy requirement is to guarantee the security of sensitive attributes values. Besides, the *SS*-attack can be thwarted by meeting the privacy requirement of Definition 7(3) that limits the frequency of ss-record in groups.

To satisfy the PPMS(3, 1/3, 1/4)-bounding, Table 3(A)–3(C) can be released for Table 1(A)–1(C). Each sanitized group incorporates at least three such individual *P* which have these properties: (a) *P* is the first time appearing in multiple releases; (b) *P* will not appear in the next release. At the same time, the new sensitive values cover the corresponding old ones according to the *OID*-bounding. Each group still contains at least three records and the frequency of each sensitive value does not exceed the threshold 1/3, even though the sets *B*,*F*,*L* and *MD* have been excluded by the adversary. Thus, the attacks in [22] can be prevented. Specially, the exclusion of set *MD* from releases has no effect on reaching privacy acquirements, so

Table 3. PPMS(3,1/3,1/4)-bounding publishes tables for Table 1.

(a) Quarter 1

CaseID	Sex	Age	ADR	Group
1	M	[46–50]	c,b	1
7	M	[46–50]	a	1
3	M	[46–50]	d	1
5	M	[46–50]	e,g	1
2	F	[21–25]	c,a	2
4	F	[21–25]	b,d	2
6	F	[21–25]	y	2

(b) Quarter 2

CaseID	Sex	Age	ADR	Group
1	M	[48–53]	c,b	1
11	M	[48–53]	a	1
12	M	[48–53]	y	1
14	M	[48–53]	h	1
16	M	[48–53]	q,d,e,p,x	1
15	*	[40–46]	x	2
18	*	[40–46]	q	2
17	*	[40–46]	y,b,c,f,g,i	2
19	*	[40–46]	o	2
3	*	[39–46]	d	3
20	*	[39–46]	x,i	3
21	*	[39–46]	j,z,v,u,k,n	3
22	*	[39–46]	q,l	3
13	*	[39–46]	h,o	3

(c) Quarter 3

CaseID	Sex	Age	ADR	Group
13	F	[39–46]	h,k	1
26	F	[39–46]	x,u	1
28	F	[39–46]	d,i	1
23	F	[39–46]	z	1
15	*	[38–46]	x	2
27	*	[38–46]	a,c	2
24	*	[38–46]	d	2
25	*	[38–46]	e,q	2

https://doi.org/10.1371/journal.pone.0250457.t003

MD-attack in example 1 can be resisted. Besides, the frequency of ss-record in groups does not exceed 1/4, the *SS*-attack (example 2) can be also thwarted.

4. Algorithm and analysis

In this section, we propose an heuristic algorithm to achieve PPMS(k, θ, α)-bounding. The related definitions and symbols are in the section 4.1. We give specific steps and illustration of the algorithm in section 4.2, meanwhile, the analysis and lemmas on which the algorithm depends are also included.

4.1 Definitions and symbols

Before stating the algorithm, we introduce new symbols as follows:

Substantial symptoms-record(ss-record): as mentioned earlier, for a record $t(t\in T)$, if t has many more symptoms/adverse drug reactions than others in T, then t is an ss-record in table T.

New-record (n-record): for a record $t(t\in T)$, t's *caseid* is its initial appearance in a released table. That is, t is a n-record in table T.

Old-record(o-record): for a record $t(t\in T)$, t's *caseid* is not the initial appearance in a released table. That is, t is an o-record in table T.

Medication discontinuation-record(md-record): for a record $t(t\in T_i)$, t's *caseid* will not appear in next table T_{i+1}, t is a md-record in T_i.

nx-record($x\in${ss, n, o, md}): if t is not x kind of record, t is nx-record. For instance, if t is not a md-record in table T, then t can be denoted as nmd-record in T.

x&y-record(x, $y\in${ss, n, o, md}): if t is x kind of record and y kind of record, then t is an x&y -record. For instance, if t is a n-record and md-record in table T, then t can be denoted as n&md-record in T.

According to the background knowledge, the adversary can derive four views for an anonymous group G. Assume the adversary knows the target individual P's QIAs values, and learns that the target is in a released table.

View 1: the adversary knows that individual P is in a specific group G, P appears for the first time in released tables and P stops medication in next quarter. We denote view 1 of group G as GV_1. GV_1 contains all the n&md-records in G.

View 2: the adversary knows that individual P is in a specific group G, P appears for the first time in released tables. We denote view 2 of group G as GV_2. GV_2 contains all the n-records in G.

View 3: the adversary knows that individual P is in a specific group G, P stops medication in next quarter. We denote view 3 of group G as GV_3. GV_3 contains all the md-records in G.

View 4: the adversary knows P is in a specific group G. We denote view 4 of group G as GV_4. Obviously, $GV_4 = G$.

$GV_x(y)$ ($x\in${1, 2, 3, 4}, $y\in${ss, n, o, md}): represents the set of all the y type of records in GV_x.

$F_{ss}(GV_x)$ ($x\in${1, 2, 3, 4}): the frequency of ss-record in GV_x.

If the adversary can disclose privacy in any one of these views, the anonymous group G is unsafe. Thus, we have to provide privacy protection in all these four views.

4.2 Algorithm of PPMS(k, θ, α)-bounding

We present an algorithm called **HA** to achieve PPMS(k, θ, α)-bounding. The overview of this algorithm is as shown in Algorithm 1.

```
Algorithm 1: HA
Input: the original dataset D_i, the previous anonymous released tables
T_pre = {T_1, T_2, ..., T_i-1}, k, θ, α
Output: anonymous released table T_i of original table D_i
1:Combine records with the same caseid into a super record;
2:for each o-record t in D_i do
3:{
4: Find a released table T_j which is the first table contains caseid of
t in T_pre;
5: Record t_pre (t_pre∈T_j) has the same caseid with t, generalize the QIA
values of t to cover that of t_pre;
6:}
7:Grouping(T_i, D_i, k, θ, α);
8:Generalization;
9:return T_i;
```

The algorithm merges records with the same caseid into super records firstly (line 1). Next, it achieves QID-bounding strategy to prevent F-attack (line 2-line 6). Then, the algorithm groups the records in current table with **procedure Grouping** (line 7). Last, the algorithm anonymizes the table and releases it (line 8-line 9). Our algorithm has made **three changes** to the PPMS_EAR [22]. We first redefine the privacy risk to satisfy θ-bounding when medication discontinuation-attack is considered (**the change 1**). The introduce and analysis of **the change 1** are as follows.

Lemma 1. To resist Medication discontinuation-attack, for any sensitive value v, the allowed largest number of v in a group G is as formula (1).

$$\eta_v(G) = \lfloor |GV_1| * \theta_v \rfloor \tag{1}$$

Proof. It is easy to know that $\eta_v(G)/|GV_1| \leq \theta_v$. Meanwhile, it is clear that $\eta_v(G)/|GV_x| \leq \theta_v$ because $|GV_x| \geq |GV_1|$ ($x \in \{2, 3, 4\}$). Thus, the frequency of v will not exceed the threshold θ_v in the four views of group G. The proof is completed.

The privacy risk is the same as that in PPMS_EAR [22] except $\eta_v(G)$, it is as formula (2).

$$PR_v(G \cup \{t\}) = \begin{cases} \dfrac{\sigma_v(G \cup \{t\})}{\eta_v(G \cup \{t\}) - \sigma_v(G \cup \{t\}) + 1}, & \sigma_v(G \cup \{t\}) \leq \eta_v(G \cup \{t\}) \\ \infty, & \sigma_v(G \cup \{t\}) > \eta_v(G \cup \{t\}) \end{cases} \tag{2}$$

$PR_v(G \cup \{t\})$ can evaluate the privacy risk caused by record t's sensitive value v after G including t. The occurrence of v in G is denoted as $\sigma_v(G)$. In fact, a record usually has multiple sensitive values in ADE data, thus the privacy risk caused by record t is as formula (3).

$$PR(G \cup \{t\}) = \begin{cases} 1 + \sum_{v \in S_t} PR_v(G \cup \{t\}), & \sigma_v(G \cup \{t\}) \leq \eta_v(G \cup \{t\}) \\ \infty, & \sigma_v(G \cup \{t\}) > \eta_v(G \cup \{t\}) \end{cases} \tag{3}$$

S_t denotes the set of all the sensitive values contained by record t.

The difference of information loss ($\triangle IL(G, t)$) between group G and group $G \cup t$ is the same as in PPMS_EAR. Thus, we can get the $\triangle PRIL(G, t)$ [22] which is as formula (4).

$$\triangle PRIL(G, t) = \triangle IL(G, t) * PR(G \cup t) \tag{4}$$

A record t has less $\triangle PRIL(G, t)$ [22] with a greater probability to be included in group G.

This use of $\triangle PRIL(G, t)$ is shown on line 10 of the **procedure grouping** which will be introduced with **the change 2**. $\triangle PRIL(G, t) = \infty$ represents that the inclusion of t will break the θ-bounding, thus t cannot be included in G when $\triangle PRIL(G, t) = \infty$. Therefore, **the change 1** is actually about the redefinition of $\triangle PRIL(G, t)$ with the consideration of *MD*-attack.

Now we illustrate **the change 2** which is included by the **procedure grouping**.

```
Procedure 1: Grouping
Input: T_i, D_i, k, θ, α
Output: T_i
/* Lines 1-24 are the steps of creating groups */
1:Randomly choose a record t from D_i;
2:D_i = D_i-t;
3:while(true)
4:{
5: Create a new empty group G;
/* Update_Group (G, t) updates G's parameters which will be used by
Jugde_α_bounding(G, t~, α). The details of the procedure Update_Group
will be introduced later */
6: G = Update_Group (G, t); //G∪t_bst, the parameters of G are updated
7: while (|GV_1|<k)
8: {
/* Jugde_α_bounding(G, t~, α) determines whether G∪t~ violates
α_bounding requirement.
Jugde_α_bounding(G, t~, α) = true indicates that G∪t~ meets the
α_bounding requirement.
Jugde_α_bounding(G, t~, α) = false indicates that the α_bounding
requirement cannot be met.
The details of the procedure Jugde_α_bounding will be introduced
later. */
9: D_i~contains all these records t~ in D_i: Jugde_α_bounding(G, t~, α) =
true;
10: Find a record t_bst (t_bst∈ D_i~)has the least △PRIL(G, t_bst) in D_i~;
11: if t_bst can not be found then // If creating groups fails
12: {
13: Add all records of G to D_i;
14: Break; //The process of creating groups is completed, jumps to the
line 19
15:}
16: G = Update_Group (G, t_bst); //G∪t_bst, the parameters of G are
updated
17: D_i = D_i-t_bst;
18:} //end while (|GV_1|<k)
19: if |GV_1|> = k then // If creating groups successes
20: T_i∪G;
21: else// If creating groups fails
22: Break;// The process of creating groups is completed, jumps to the
line 25
23: Choose a record t which is farthest from t_bst;
24:}//end while(true)
/*Lines 25-30 are the steps of processing the remaining records*/
25:for each t∈D_i do
26:{
27: Find a group G from T_i which has the least △PRIL(G, t);
28: G∪t;
29: D_i = D_i-t;
30:}
```

In the **procedure grouping**, to resist medication discontinuation-attack, for each group G, the $|GV_1|$ of G should be no less than k (**the change 2**, line 7- line 18). When $|GV_1| = k$, group G is completed; otherwise, the grouping of G will continue. Thus, the k-bounding can be guaranteed. Besides, the procedure processes the remaining records after the completion of grouping step (line 25-line 30).

The third change is also in **procedure grouping**. We should verify if the α-bounding can be satisfied. However, before G is completed, $|GV_x|$ ($x \in \{2, 3, 4\}$) cannot be known. Thus, we find a way to verify α-bounding in the group process.

Lemma 2. $F_{ss}(GV_2)$ is no more than $|GV_2(ss)| / (|GV_2 - GV_2(n\&nmd\&nss)|)$.

Proof. It is easy to know that $|GV_2 - GV_2(n\&nmd\&nss)| \leq |GV_2|$. Thus, $|GV_2(ss)| / (|GV_2 - GV_2(n\&nmd\&nss)|) \geq |GV_2(ss)| / |GV_2| = F_{ss}(GV_2)$.

Similarly, we can get:

$F_{ss}(GV_3) \leq |GV_3(ss)| / (|GV_3 - GV_3(o\&nss)|)$.

$F_{ss}(GV_4) \leq |GV_4(ss)| / (|GV_4 - GV_4(o\&nss) - GV_4(n\&nmd\&nss)|)$.

According to the above observations, it is easy to know that GV_2 meets the a- bounding requirement when $a \geq |GV_2(ss)| / (|GV_2 - GV_2(n\&nmd\&nss)|)$. The similar conclusions can be got for GV_3 and GV_4. According to these conclusions, we can verify α-bounding through **procedure Jugde_α_bounding** and **procedure Update_Group.** These two procedures are shown below.

```
Procedure 2: Jugde_α_bounding
Input: G, t
Output: true or false
G is completed in one of these conditions:
In grouping step, the grouping of G is completed;
In the step of processing remaining record, a remaining record is
added to G.
G.MSAVR1 represents GV1(SS) after group G is completed;
G.MSAVR2 represents GV2(SS) after group G is completed;
G.MSAVR3 represents GV3(SS) after group G is completed;
G.MSAVR4 represents GV4(SS) after group G is completed;
G.N1 represents |GV1 |after group G is completed;
G.N2 represents |GV2- GV2(n&nmd&nss) |after group G is completed;
G.N3 represents |GV3- GV3(o&nss) |after group G is completed;
G.N4 represents |GV4- GV4(o&nss)- GV4(n&nmd&nss) |after group G is
completed;
1:When G is created without records, G.MSAVRx = 0, G.Nx = k(x∈{1, 2, 3,
4}).
2:G = Update_Group (G, t);
3:if (G.MSAVR1 / G.N1 >α || G.MSAVR2 / G.N2 >α || G.MSAVR3 /G.N3 >α ||
G.MSAVR4 / G.N4 >α)
4: return false;
5:else
6: return true;
```

The **procedure Jugde_α_bounding** determines whether the group G violates α_bounding requirement after incorporating the record t. According to **procedure 1**, each group contains at least k n&md-records when the creation of it is completed. The set R_G is used to represent these k n&md-records of G. It is clear that there are $R_G \subseteq GV_1$, $R_G \subseteq (GV_2 - GV_2(n\&nmd\&nss))$, $R_G \subseteq (GV_3 - GV_3(o\&nss))$ and $R_G \subseteq (GV_4 - GV_4(o\&nss) - GV_4(n\&nmd\&nss))$, when the creation of G is completed. Thus, we reserve the positions for the above k n&md-records in GV_1, GV_2-GV_2(n&nmd&nss), GV_3- GV_3(o&nss) and GV_4-GV_4(o&nss)-GV_4(n&nmd&nss), respectively. The initial value of $G.N_x$ ($x \in \{1, 2, 3, 4\}$) is k (line 1). Besides, we assume that all the above k n&md-records are nss-records at the beginning of the creation of G, so the initial value of G.

$MSAVR_x$ is *0* (x∈{1, 2, 3, 4})(line 1). After updating $G.MSAVR_x$ and $G.N_x$ (line 2), the **procedure Jugde_α_bounding** verifies α_bounding: if all the four views of group *G* can meet the α_bounding requirement, then the α_bounding can be met by *G* (line 3-line 6). The main function of the **procedure Update_Group** is to update $G.MSAVR_x$ and $G.N_x$ after incorporating the record *t*.

```
Procedure 3: Update_Group
Input: G, t
Output: G
1:G = G∪t
2:if t is n&md-record then
3:{
4: if t is ss-record then
5: {
6: G.MSAVR1++; G.MSAVR2++; G.MSAVR3++; G.MSAVR4++;
7: if t is a remaining record then
8: {
/* t∉RG, t∈GVx(x∈{1, 2, 3, 4}), t∈(GV2-GV2(n&nmd&nss)), t∈(GV3-
GV3(o&nss)) and t∈(GV4- GV4(o&nss)- GV4(n&nmd&nss)) */
9: G.N1++; G.N2++; G.N3++; G.N4++;
10:}
11:}
12}
13:else if t is n&nmd-record then
14:{
15: if t is ss-record then
16: {
/* t∉RG, t∈GV2, t∈GV4, t∈(GV2- GV2(n&nmd&nss)), t∈(GV4- GV4(o&nss)-
GV4(n&nmd&nss))*/
17: G.MSAVR2++; G.MSAVR4++; G.N2++; G.N4++;
18:}
19:}
20:else if t is o&md-record then
21:{
22: if t is ss-record then
23: {
/* t∉RG, t∈GV3, t∈GV4, t∈(GV3- GV3(o&nss)), t∈(GV4- GV4(o&nss)-
GV4(n&nmd&nss))*/
24: G.MSAVR3++; G.MSAVR4++; G.N3++; G.N4++;
25:}
26:}
27:else
28:{
29: if t is ss-record then
30: {
/* t∉RG, t∈GV4, t∈(GV4- GV4(o&nss)- GV4(n&nmd&nss))*/
31: G.MSAVR4++; G.N4++;
32:}
33:}
34:return G;
```

For various types of records, the **procedure Update_Group** updates $G.MSAVR_x$ and $G.N_x$ (x∈{1,2,3,4}). Now we take the lines 2–12 as an example. These lines consider $G.MSAVR_x$ and $G.N_x$ when *t* is a n&md-record. It is clear that there is $t \in GV_x$(x∈{1,2,3,4}) when *t* is a n&md-record. Thus, $G.MSAVR_x$ is updated if *t* is an ss-record (line 6). On the other hand, the positions of *t* have been already reserved in GV_1, GV_2-GV_2(n&nmd&nss), GV_3-GV_3(o&nss) and GV_4-GV_4(o&nss)-GV_4(n&nmd&nss), when *t* is not a remaining record. Therefore, $G.N_x$ (x∈

{1,2,3,4}) remains unchanged when t is not a remaining record. However, there are $t \notin R_G$, $t \in GV_1$, $t \in (GV_2\text{-}GV_2(\text{n\&nmd\&nss}))$, $t \in (GV_3\text{-}GV_3(\text{o\&nss}))$ and $t \in (GV_4\text{-}GV_4(\text{o\&nss})\text{-}GV_4(\text{n\&nmd\&nss}))$ when t is remaining record and ss-record. Thus, $G.N_x$ ($x \in \{1,2,3,4\}$) is updated (line 9). Similarly, the other parts of **procedure Update_Group** update $G.MSAVR_x$ and $G.N_x$ according to the type of t.

Our algorithm quits predetermining the maximum number of ss-records, so it is more flexible. The experimental results also show that our heuristic algorithm can maintain the usability of released tables when has more stringent privacy requirement.

5. Experimental results and analysis

In this section, we compare our method (**PPMS(k, θ, α)-bounding** achieved by **HA**) with **PPMS_Ear** [22]. We implement both the methods with Microsoft Visual C++ 2015. All the experiments are conducted on a PC with Intel Core 2.60 GHz CPU and 8 GB main memory, running the Microsoft Windows 10 operating system.

We analyze the methods from security and information loss. The 14 recent datasets are chosen from FEARS of FDA:2014Q3-2017Q4. The quasi-identifiers (QIAs) and sensitive attributes(SAs) are the same as in [22], QIAs = {Weight, Age, Gender}, SAs = {INDI_PT, PT}. To define ss-record, we calculate the average count $AVG_{\text{INDI\_PT}}$/AVG_{PT} of INDI_PT/PT values and the corresponding standard deviation $SD_{\text{INDI\_PT}}$/ SD_{PT}. We define record t is a **ss-record**, if t satisfies one of these conditions:

1. The count of INDI_PT values of t is no less than $AVG_{\text{INDI\_PT}}+SD_{\text{INDI\_PT}}$;
2. The count of PT values of t is no less than $AVG_{\text{PT}}+SD_{\text{PT}}$.

5.1 Security

Dangerous Identity Ratio(DIR) [21–22] and Dangerous Sensitivity Ratio (DSR) [21–22] are used to evaluate the security of publishing methods. We call a group as a dangerous identity group (DIG) if the number of records in the group is less than threshold k. If a group contains at least one sensitive value v_i whose frequency is higher than its threshold θ_i, we call it as a dangerous sensitivity group (DSG). DIR/DSR represents the ratio of DIG/DSG in all anonymous groups.

For measuring the ability to resist substantial symptoms-attack, we define the substantial symptoms group ratio (SSGR). If the frequency of ss-record in a group is higher than threshold α, we call the group as substantial symptoms group(SSG). Similar to DIR and DSR, SSGR represents the ratio of SSG in all anonymous groups.

As shown in Figs 1 and 2, the DIR and DSR of PPMS_Ear are both greater than 0, because PPMS_Ear has not taken the medication discontinuation-attack into consideration. The DIR and DSR are even greater than 10% in some released tables of PPMS_Ear. An adversary can compromise the privacy requirements in the released tables of PPMS_Ear which is vulnerable to the medication discontinuation-attack. lMeanwhile, PPMS(k, θ, α)-bounding has considered that an adversary may disclose privacy through information about medication discontinuation. Therefore, PPMS(k, θ, α)-bounding can avoid Medication discontinuation-attack, DIR and DSR are both 0.

The SSGR of the two methods is shown in Fig 3. We can see that the SSGR of PPMS(k, θ, α)-bounding is 0 because our method can thwart *SS*-attack with limiting the frequencies of ss-records. However, PPMS_Ear cannot resist *SS*-attack, hence its SSGR in some released tables is greater than 10%. The settings of θ($\theta = 0.4$, $\theta =$ bf) are omitted, because they generate similar results.

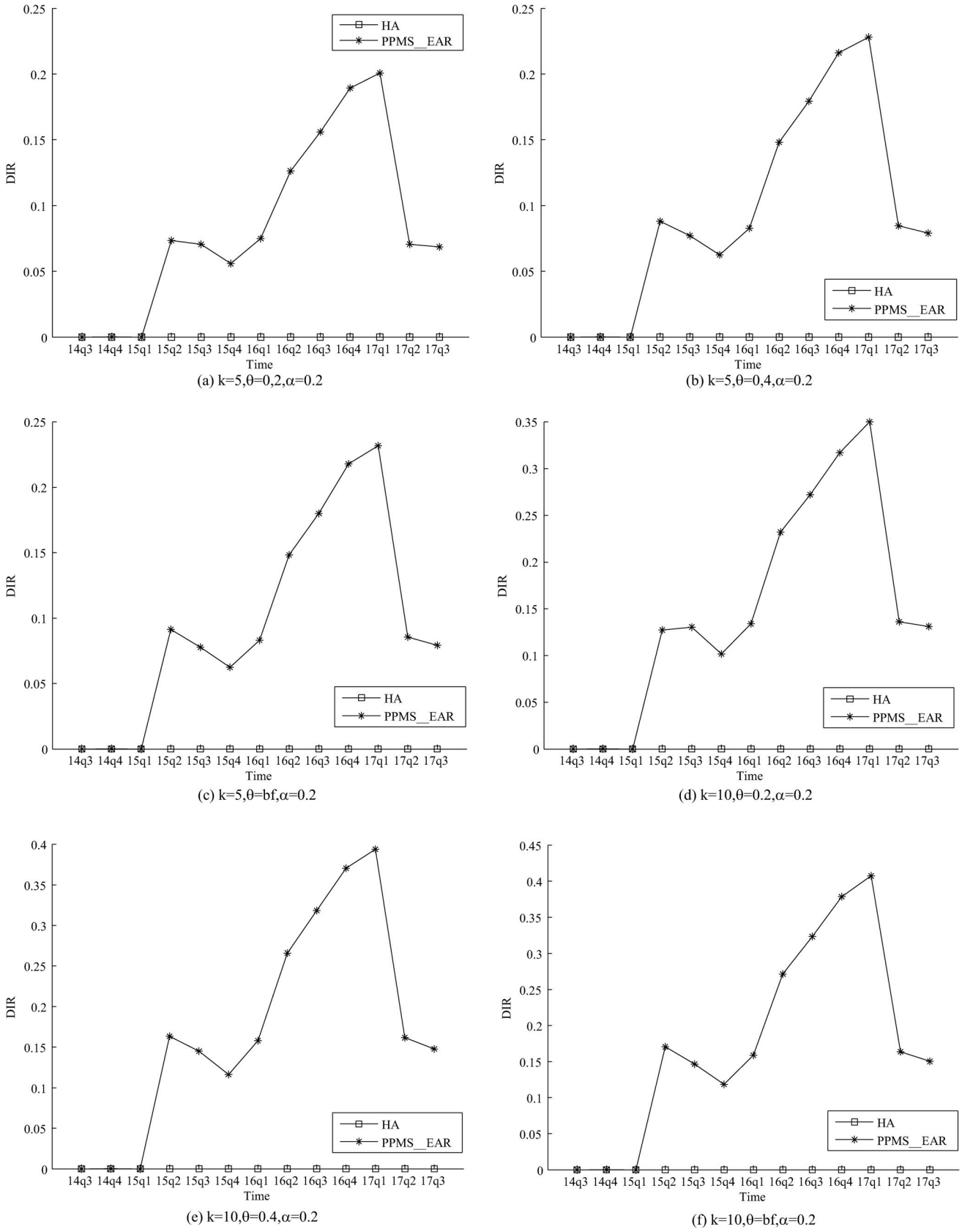
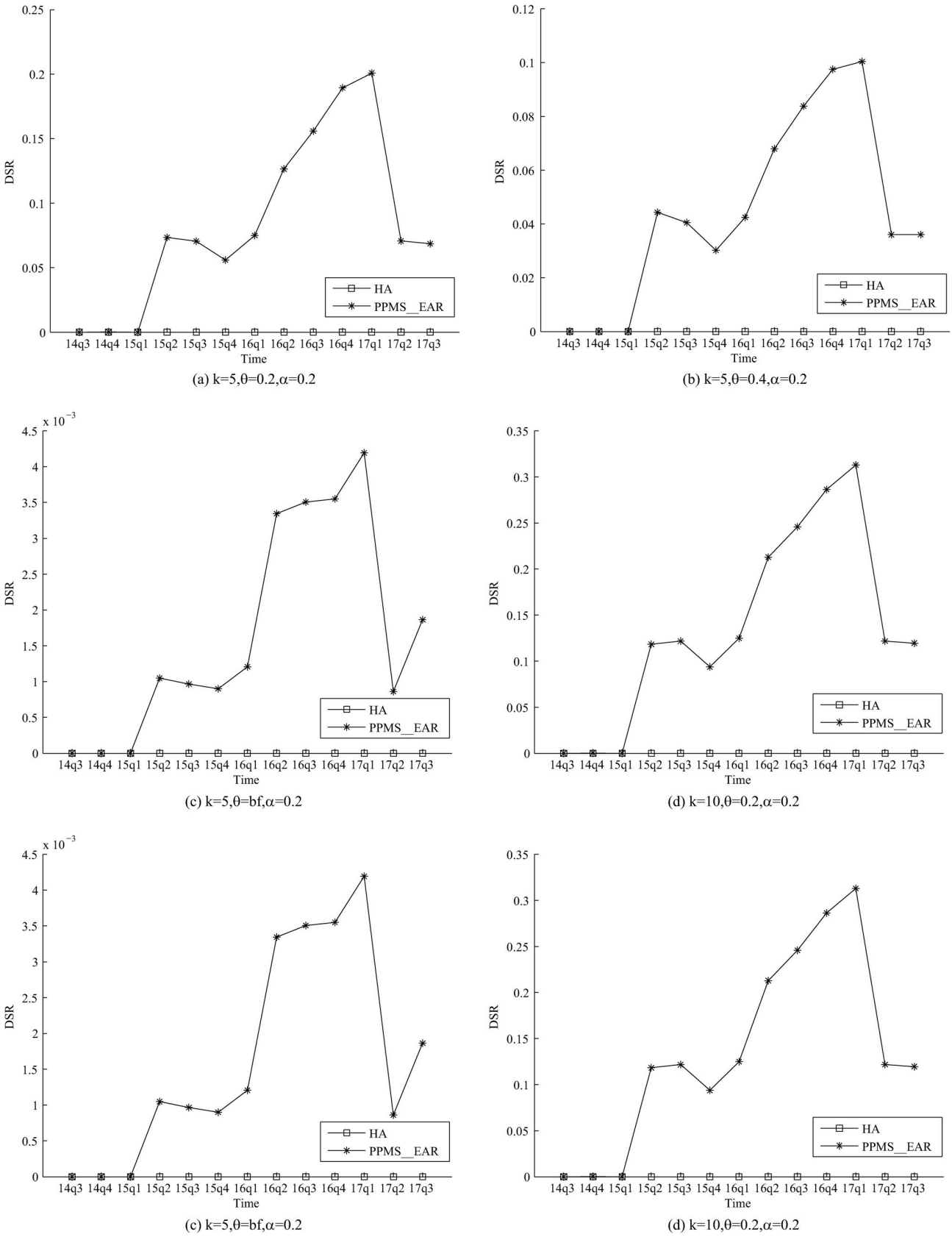


Fig 1. Evaluation on DIR of two methods.

https://doi.org/10.1371/journal.pone.0250457.g001



(a) k=5,θ=0.2,α=0.2

(b) k=5,θ=0.4,α=0.2

(c) k=5,θ=bf,α=0.2

(d) k=10,θ=0.2,α=0.2

(c) k=5,θ=bf,α=0.2

(d) k=10,θ=0.2,α=0.2

Fig 2. Evaluation on DSR of two methods.

https://doi.org/10.1371/journal.pone.0250457.g002

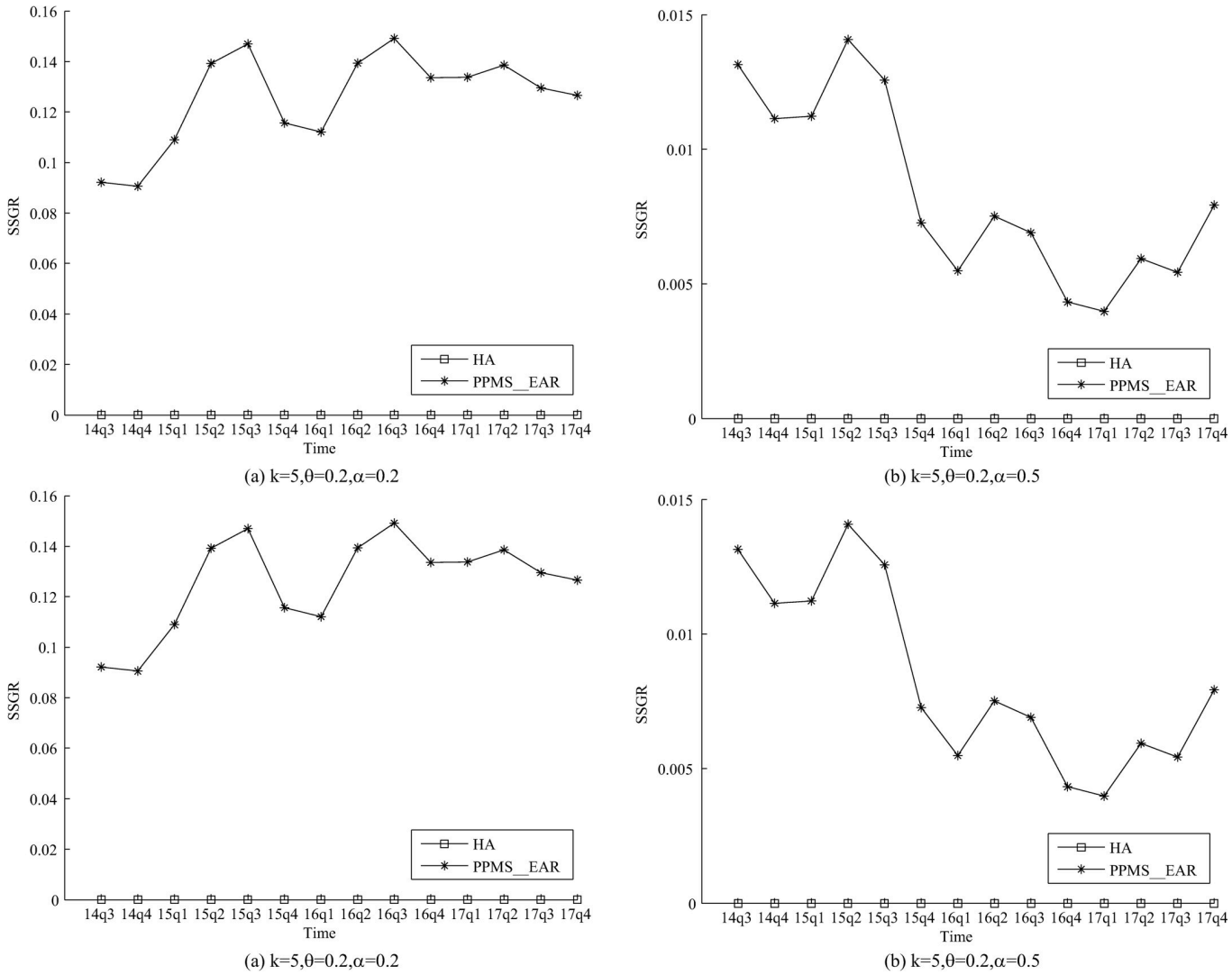


Fig 3. Evaluation on SSGR of two methods.

https://doi.org/10.1371/journal.pone.0250457.g003

5.2 Information loss

Normalized Information Loss(NIL) [21–22] is used to evaluate the information usability. As shown in Fig 4, we can see that the NIL of PPMS(k, θ, α)-bounding is very close to PPMS_Ear. Our analysis result is that **procedure Jugde_α_bounding** achieves α-bounding through a heuristic method, and it is easier for records to be incorporated by groups in this method. The **HA** estimates the frequencies of ss-records to guarantee α-bounding when grouping. Compare with predetermining the maximum number of ss-records in groups, this heuristic method can "accommodate" more ss-records in each anonymous group while the privacy requirement is not compromised. Therefore, our method can have similar information loss with PPMS_Ear even if the privacy requirements of ours are more stringent.

6. Discussion

The tradeoff between privacy and utility is the focus in data publishing. The traditional periodical data publishing mechanisms [9–14] are not suitable for the SRS data due to its some special

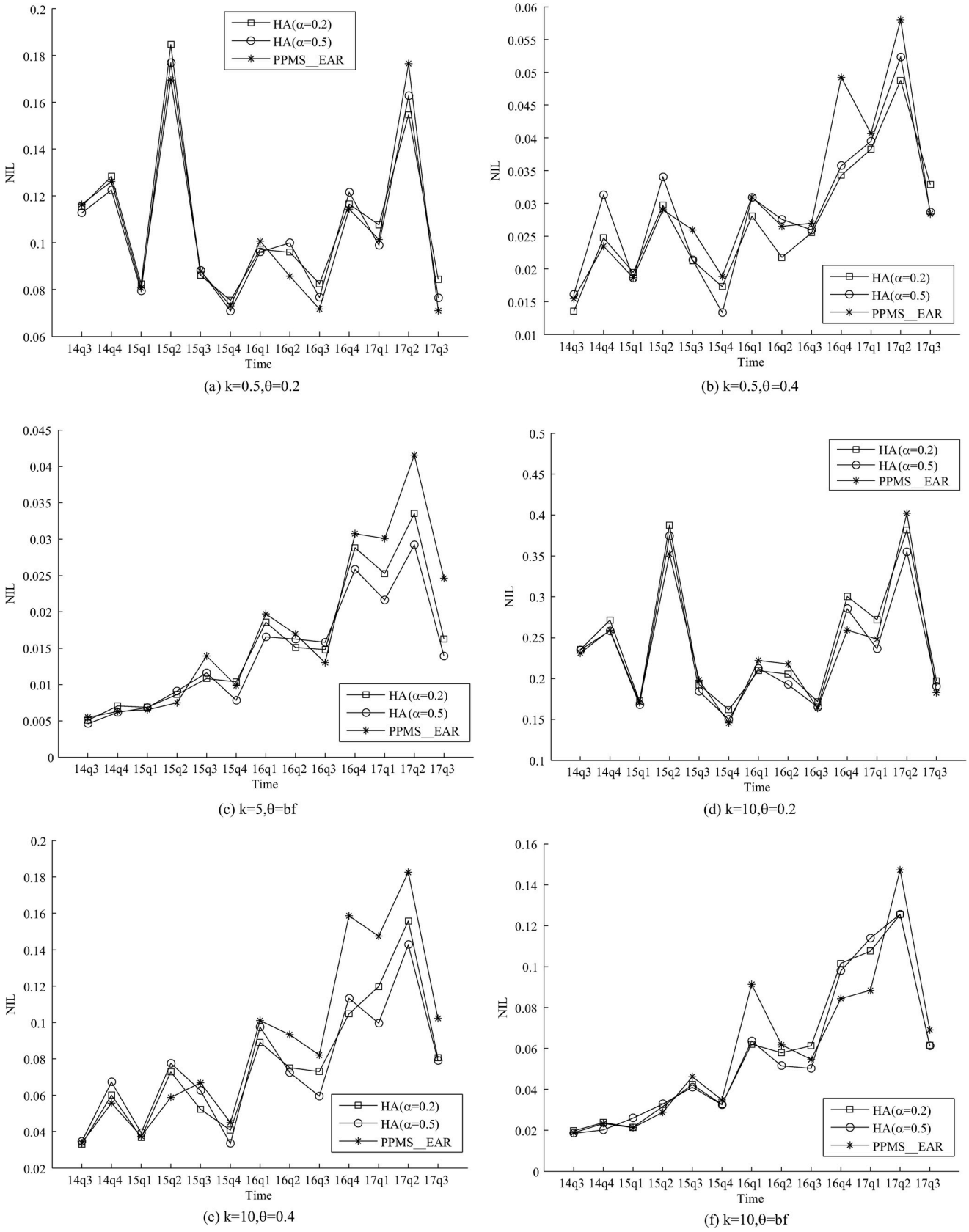


Fig 4. Evaluation on NIL of two methods.

https://doi.org/10.1371/journal.pone.0250457.g004

features. The existing SRS data publishing methods [22–25] had considered these types of attacks in the scenario of SRS data publishing: *B*-attack, *F*-attack and *L*-attack, trying to guarantee the information usability of the released tables. However, the SRS data has some special features, such as identity preservation and multivalued sensitive attributes, which make it more vulnerable to security threats than the other types of data. We discover and define two new attacks in this paper: *MD*- attack and *SS*-attack, and find that the existing SRS data publishing methods cannot resist them. In order to solve these problems, we present a new SRS data publishing model and the corresponding heuristic algorithm in this paper. We consider these attacks in the evaluations of DIR and DSR: *MD*-Attack, *B*-attack, *F*-attack and *L*-attack, the related experimental results (Figs 1 and 2) show that all these attacks can be resisted by the proposed method. The evaluation of SSGR is used to analyze the *SS*-attack, and the corresponding results (Fig 3) also demonstrate that the proposed method can thwart this type of attack. Thus, the results of security evaluations exhibit that the proposed method can defend against the various known attacks. For the information usability, the related experimental results (Fig 4) demonstrate that the information loss degree of the proposed method is similar to the existing one. The results of information usability evaluation suggest that the proposed heuristic algorithm is an effective way to decrease the information loss when the anonymous standard becomes more stringent.

According to the above experimental results of security and usability, we can know that the proposed method can provide better protection than the exiting SRS publishing methods and the guaranteed information usability can be provided by our method. Compared with the existing SRS publishing methods, our method can achieve a better balance between privacy security and information usability.

7. Conclusion

In this paper, we consider medication discontinuation and substantial symptoms in periodical ADEs data publishing, and propose a new periodical ADEs data publishing method, which can resist medication discontinuation-attack and substantial symptoms attack. The experimental results also show that our method can protect against various known attacks, and can enhance the security based on PPMS_Ear. Besides, comparing with PPMS_Ear, our method does not have an obvious increase in information loss.

Several directions for future work are also initiated by this work. First, it would be interesting to study personalized anonymity [26] in SRSs data publishing. This technique enables individuals to specify privacy levels to their own sensitive information, in order to yield a better tradeoff between privacy security and information utility. Second, it would be exciting to extend the proposed method to be applicable for big data publication [27]. Research towards this direction may discover effective parallel algorithms with guaranteed privacy security and information usability.

Acknowledgments

Thanks go to Mike Brewes of Nipissing University, Canada for his assistance in proofreading this article.

Author Contributions

Conceptualization: Wei Huang.

Data curation: Haibin Zhu.

Formal analysis: Haibin Zhu.

Funding acquisition: Wei Huang, Tong Yi.

Investigation: Wenqian Shang.

Methodology: Wei Huang, Tong Yi.

Project administration: Weiguo Lin.

Resources: Wenqian Shang.

Software: Wei Huang, Tong Yi.

Supervision: Weiguo Lin.

Validation: Tong Yi.

Visualization: Wenqian Shang.

Writing – original draft: Wei Huang, Tong Yi.

Writing – review & editing: Tong Yi, Haibin Zhu.

References

1. FDA Adverse Event Reporting System (FAERS), https://open.fda.gov/data/faers.
2. The Yellow Card Scheme, http://yellowcard.mhra.gov.uk.
3. Sweeney L. k-anonymity: a model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.2002; 10(5): 557–570. https://doi.org/10.1142/S0218488502001648
4. Abdalaal A, Nergiz ME, Saygin Y. Privacy-preserving publishing of opinion polls. Computers & Security. 2013; 37: 143–154. https://doi.org/ 10.1016/ j.cose. 2013. 03.008
5. Lee H, Kim S, Kim J W, Chung Y D. Utility-preserving anonymization for health data publishing. Bmc Medical Informatics & Decision Making.2017; 17, 12 pages. https://doi.org/10.1186/s12911-017-0499-0 PMID: 28693480
6. Li T C, Li N H, Zhang J. Slicing: A new approach for privacy preserving data publishing. IEEE Transactions on knowledge and data engineering. 2012; 24(3): 561–574. https://doi.org/10.1109/tkde.2010.236
7. Wang H, Liu R. Hiding outliers into crowd: Privacy-preserving data publishing with outliers. Data & Knowledge Engineering. 2015; 100: 94–115. https://doi.org/10.1016/j.datak.2015.06.012
8. Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam, M. L-diversity: privacy beyond kanonymity. International Conference on Data Engineering (ICDE '06). Atlanta, Georgia,2006;24–35. https://doi.org/10.1109/icde.2006.1
9. Fung B C M, Wang K, Fu A W C, Pei J. Anonymity for continuous data publishing. Proceedings of the 11th international conference on Extending database technology: Advances in database technology. Nantes, France, 2008;264–275. https://doi.org/10.1145/1353343.1353378
10. Xiao X, Tao Y F. m-invariance: Towards privacy preserving re-publication of dynamic datasets. ACM SIGMOD International Conference on Management of Data. Beijing, China, 2007;689–700. https://doi.org/10.1145/1247480.1247556
11. Pei J, Xu J, Wang Z, Wang W, Wang K. Maintaining K-Anonymity against Incremental Updates. International Conference on Scientific and Statistical Database Management. Bamff, AB, Canada, 2007;1–10. https://doi.org/ 10.1109/ ssdbm.2007.16
12. Li F, Zhou S. Challenging more updates: towards anonymous republication of fully dynamic datasets. Computing Research Repository,abs/0806.4703, 2008.
13. Anjum A, Raschia G. Anonymizing sequential releases under arbitrary updates. EDBT/16thICDT 2013 Workshop on Privacy and Anonymity in the Information Society. Genoa, Italy, 2013;145–154. https://doi.org/10.1145/2457317.2457342
14. Bewong M, Liu J, Liu L, Li J. Privacy preserving serial publication of transactional data. Information Systems.2019; 82:53–70. https://doi.org/10.1016/j.is.2019.01.001
15. Wang D, Xu Z. Impact of inaccurate data on Differential Privacy. Computers and Security.2019; 82:68–79. https://doi.org/10.1016/j.cose.2018.12.007

16. Liu F. Generalized Gaussian Mechanism for Differential Privacy. IEEE Transactions on Knowledge and Data Engineering.2019; 31(4):747–756. https://doi.org/10.1109/tkde.2018.2845388

17. Shin H, Kim S, Shin J, Xiao X. Privacy Enhanced Matrix Factorization for Recommendation with Local Differential Privacy. IEEE Transactions on Knowledge and Data Engineering.2018; 30(9):1770–1782. https://doi.org/ 10.1109/ tkde.2018.2805356

18. Wang P, Zhang H. Differential privacy for sparse classification learning. Neurocomputing. 2020; 375:91–101. https://doi.org/10.1016/ j.neucom. 2019. 09. 020

19. Bewong M, Liu J, Liu L, Li J, Choo K K R. A relative privacy model for effective privacy preservation in transactional data. Concurrency & Computation: Practice & Experience.2019; 31(23). https://doi.org/10.1002/cpe.4923

20. Lin W Y, Yang D C. On privacy-preserving publishing of spontaneous ADE reporting data. IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Shanghai, China, 2013. https://doi.org/10.1109/bibm.2013.6732760

21. Lin W Y, Yang D C, Wang J T. Privacy preserving data anonymization of spontaneous ADE reporting system dataset. BMC Medical Informatics and Decision Making. 2016; 16(suppl 1):21–35. https://doi.org/10.1186/s12911-016-0293-4 PMID: 27454754

22. Wang J T, Lin W Y. Privacy Preserving Anonymity for Periodical SRS Data Publishing. International Conference on Data Engineering. San Diego, California, USA, 2017;1344–1355. https://doi.org/10.1109/icde.2017.176

23. Hsiao M H, Lin W Y, Hsu K Y, Shen Z X. On Anonymizing Medical Microdata with Large-Scale Missing Values—A Case Study with the FAERS Dataset *. 41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2019;6505–6508. https://doi.org/10.1109/embc.2019.8857025

24. Lin W, Hsu K, Shen Z. Privacy-Preserving SRS Data Anonymization by Incorporating Missing Values. Conference on Technologies and Applications of Artificial Intelligence (TAAI). Taichung, 2018; 106–109. https://doi.org/ 10.1109/ taai.2018.00032

25. Cui Z, Zhang L, Wu B, Zhao Z, Mei Z, Wu Z. Efficient Q-Value Zero-Leakage Protection Scheme in SRS Regularly Publishing Private Data. Technical Gazette.2019; 26(3):695–702. https://doi.org/10.17559/tv-20181218005415

26. Wang J, Cai Z, Yu J. Achieving Personalized k-Anonymity-Based Content Privacy for Autonomous Vehicles in CPS. IEEE Transactions on Industrial Informatics. 2020; 16(6):4242–4251. https://doi.org/10.1109/tii.2019.2950057

27. Zakerzadeh H, Aggarwal C C, Barker K. Privacy-preserving big data publishing. International Conference on Scientific and Statistical Database Management. 2015. https://doi.org/10.1145/2791347.2791380