This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Distributed Architecture with Double-Phase Microaggregation for the Private Sharing of Biomedical Data in Mobile Health

Agusti Solanas, *Member, IEEE,* Antoni Martínez-Ballesté, *Member, IEEE,* and Josep M. Mateo-Sanz

***Abstract*—In this article we present the concept of double-phase microaggregation as an improvement of classical microaggregation for the protection of privacy in distributed scenarios without fully trusted parties.**

We apply this new concept in the context of mobile health and we show that a distributed architecture consisting of patients and several intermediate entities can apply it to protect the privacy of patients, whose data are released to third parties for secondary use.

After recalling some fundamental concepts of statistical disclosure control and microaggregation, we detail the distributed architecture that allows the private gathering, storage and sharing of biomedical data. We show that double-phase multivariate microaggregation properly fits the needs for privacy preservation of biomedical data in the distributed context of mobile health. Moreover, we show that double-phase microaggregation performs similarly to classical microaggregation in terms of information loss, disclosure risk and correlation preservation, whilst avoiding the limitations of a centralised approach.

***Index Terms*—Privacy protection, Distributed environments, Microaggregation, Mobile health**

I. Introduction

DUE to the recent advances in information and communication technologies the gathering, storage and sharing of data are becoming simpler and faster than ever. In this regard, one of the most significant breakthroughs is the use of mobile devices (e.g., cell phones) to monitor patients. Mobile communications are experiencing a tremendous development that leads to new ways of providing healthcare services, the so-called mobile health (m-health). m-Health affects the way we understand healthcare services in three main aspects: (a) m-Health simplifies the access to classical and new services. The ubiquity of mobile devices allows services to be accessed everywhere, anytime. As a result, data could be collected more easily regardless of the location of patients; (b) m-Health is patient-oriented. Patients play a key role in an m-health service, because in most cases they are responsible for the remote control of the service and; (c) m-Health is personalized. Patients receive customized services that fit their specific needs. m-Health and, in general, e-health significantly contribute to the efficiency and immediacy of healthcare services and, as a result, to the treatment quality received by patients. Last but not least, the generalisation of m-health can drastically increase the amount of data (and biomedical data in particular) that can be collected, stored and analysed.

Thanks to commonly used mobile devices such as cell phones attached to the proper gadgets, it is possible to monitor variables such as the heartbeat rate or the blood pressure of patients almost constantly regardless of their location. This new ability of gathering huge amounts of personal, highly sensitive data opens the door to new studies that might use diverse information coming from multiple sources.

Notwithstanding the easy gathering of data does not directly translate into the easy access to and use of those data by researchers. In fact, according to privacy legislation, patients must give their consent to allow researchers to study their data for a secondary use. Unfortunately, it has been observed that requiring explicit consent for the participation in different forms of health research can negatively affect the process and outcomes of the very research [6]. In fact, people who consent tend to be different from those who decline. Hence, the results of the studies can be biased. With the aim to alleviate this problem, many research ethics boards will waive the consent requirement if the data used are deemed to be de-identified. As a result, there is an increasing need for methods that allow the gathering and de-identification of biomedical data for secondary use.

A. Contribution and plan of the article

In this article we propose the new concept of double-phase microaggregation as an evolution of classical microaggregation for distributed contexts in which there are no fully trusted parties. We take the distributed scenario provided by m-Helath and we describe a distributed privacy-aware architecture to collect, store, protect and release data from patients. Due to the de-identification performed by a double-phase microaggregation process, these data can be accessed by research centres for secondary use without the need for explicit patient consent.

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org

A. Solanas is with the UNESCO Chair in Data Privacy and the Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Catalonia, Spain, e-mail: agusti.solanas@urv.cat.

A. Martínez-Ballesté is with the UNESCO Chair in Data Privacy and the Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Catalonia, Spain, e-mail: antoni.martinez@urv.cat.

J.M. Mateo-Sanz is with the Area of Statistics and Operational Research, Universitat Rovira i Virgili, Catalonia, Spain, e-mail: josepmaria.mateo@urv.cat.

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

The rest of the article is organized as follows: In Section II we recall some basic definitions and concepts related to statistical disclosure control, microaggregation and public key cryptography. Next, in Section III we justify the need for a distributed architecture that is described and analysed in detail. In Section IV, we analyse the effect of a double-phase microaggregation in terms of information loss and disclosure risk over a biomedical dataset. Finally, the article concludes in Section V with some final comments and pointing out some further research lines.

II. Background

A. Statistical disclosure control

Protecting individual privacy is paramount for many institutions, namely statistical agencies, healthcare centres, Internet companies, manufacturers, etc. Many efforts have been devoted to develop techniques that assure a given degree of privacy to the people, whose data are collected and shared. Currently, these efforts come from very diverse fields of knowledge, namely cryptography, statistics, artificial intelligence, etc. Among all these disciplines, statistical disclosure control (SDC) [18], [19], [10] was the first to consider the problem; initially on tabular data, and later on microdata (i.e. data from individuals).

In this article we focus on the protection of individual privacy by means of protecting their microdata. By doing so, we aim at avoiding the re-identification of individuals through their released microdata. To achieve this goal microdata sets have to be properly modified prior to their publication. The degree of modification varies between two extremes: (i) encrypting the microdata and, (ii) leaving the microdata intact.

In the first extreme, the protection is perfect (i.e. only the owner of a decryption key can see the data), however, the utility of the data is almost nonexistent because the encrypted microdata can be hardly studied or analysed by others than the owner of the decryption key. Although there exist fully homomorphic encryption schemes that allow computations on encrypted data, they are still not fully practical and require a significant computation effort, that is not affordable to most services yet [9]. In the other extreme, the microdata are totally useful (i.e. all their information remains intact), however, the privacy of the individuals is endangered because sensitive private data can be seen without limitation.

SDC methods aim at distorting the original microdata sets to protect individuals privacy and avoid their re-identification while maintaining some of the statistical properties of the data and minimising the information loss as much as possible.

B. Basic definitions and concepts

Next we recall some basic definitions related to the field of statistical disclosure control that are used throughout the article:

- **Microdata.** In opposition to macrodata, that refers to large aggregates of information generally represented in tables, microdata refers to individual data such as the social security number (SSN), age, ethnicity, height, income, etc. that are represented with records. In our example shown in Table I, each row represents a microdata record.
- **Microdata set.** A microdata set is the union of microdata records sharing the same attributes. Thus, a microdata set is understood as a two-dimensional matrix in which rows represent individual data and columns represent specific attributes. Table I is an example of a microdata set with six attributes belonging to four individuals.
- **Identifiers.** Those attributes in a microdata set that point out to a unique individual are called identifiers. In our example the social security number (SSN) is an identifier. (This attributes are deleted before releasing microdata for public use).
- **Quasi-identifiers.** Those attributes containing information about an individual that, when taken individually, do not identify him/her. Note that, the combination of quasi-identifiers might lead to the identification of a unique individual. Examples of this kind of attribute in Table I could be zip code, weight and height (e.g. a very tall person in a small village could be easily identified).
- **Confidential outcome attributes**. Those attributes that have sensitive information like religion, salary, health condition, etc. In our example the existence of a given disease is a confidential outcome attribute.

TABLE I
EXAMPLE OF MICRODATA SET WITH 4 RECORDS AND 6 ATTRIBUTES.

Social Security Number	**Zip Code**	**Height (cm)**	**Weight (kg)**	**Av. heart-beat rate (bpm)**	**Disease**
123-23-1234	00501	169	77	80	NO
111-90-9087	04032	220	130	65	YES
881-00-2355	55802	155	50	92	NO
570-35-8104	90501	172	68	78	NO

C. The microaggregation problem

To protect a microdata set it is essential to remove all identifiers before releasing the data. Thus, attributes like the SSN or the full name should be deleted. However, removing identifiers is not enough protection because the combination of quasi-identifiers might lead to the re-identification of individuals. For example, consider the case of two attributes (Occupation and Town). If these attributes were kept untouched in the microdata set, it would be very simple to identify the doctor of a little village because, in general, there is a single doctor in a small village and everybody knows him/her. A straightforward solution to this problem is to remove all quasi-identifiers. However, due to the fact that almost every attribute could

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

be considered a quasi-identifier, their removal results in a useless/empty microdata set.

With the aim to solve this problem, a variety of techniques have been proposed, namely noise addition [1], [4], rank swapping [3], [8], generalization and deletion, microaggregation, etc. Microaggregation is a technique that protects the privacy of individuals by aggregating similar microdata records and producing microaggregated data sets satisfying the property of k-anonymity [13], [14], [17]. Given a microdata set $\mathcal{M}$, it is said that it is k-anonymous if for every record r_i there are, at least, other $k-1$ identical, undistinguishable records.

Microaggregating a microdata set can be understood as a clustering problem with constraints on the cardinality of the clusters. Given a privacy parameter k and a microdata set $\mathcal{M}$ with n records, the microaggregation problem consists in obtaining a partition of the microdata set, so that each part in the partition has at least k records and the within-part homogeneity is maximized (i.e. records in the same part/cluster are as similar as possible). The resulting partition is called a k-partition. Once the k-partition is obtained, a microaggregated data set is generated by replacing each original record by the average record of the part to which it belongs. Every record in the microaggregated data set has, at least, $k-1$ indistinguishable records.

The microaggregation problem is known to be NP-Hard in the multivariate case. Hence, several heuristics have been proposed to cope with it. Many classifications can be applied to microaggregation techniques, however, for the sake of brevity we classify them in two classes: (i) fixed-size microaggregation and, (ii) variable-size microaggregation. The former builds clusters of a fixed number of elements k, while the latter generates clusters with a variable number of records in the range $[k, 2k-1]$. A classical, well-known fixed-size microaggregation method is the maximum distance to average vector (MDAV) [5]. Regarding variable-size methods, amongst other alternative we find the variable maximum distance to average vector (V-MDAV) [15] and some newer approaches that mix artificial intelligence techniques like genetic algorithms (GA) with classical methods like MDAV or V-MDAV [16]. For a complete introduction to SDC and microaggregation methods refer to [19].

D. Basics of public key cryptography

Public key cryptography, also known as asymmetric cryptography, refers to cryptographic systems that require a pair of different keys to operate: one key (the public key) K_P is used to encrypt messages and, the other key (the private key) K_{PR} is used to decrypt them[1]. The public key can be released so as to allow everyone to know it, while the private key is kept secret and is only known by its owner. In this way, private communications are possible because everyone can use the known public keys to encrypt messages and send them to the owner of the corresponding private key, who will be the only one able to decrypt and read the messages. Given a message m and a pair of keys $\{K_P, K_{PR}\}$, the sender encrypts m by using K_P,

$$\mathcal{ENC}_{K_P}(m)$$

and the receiver uses K_{PR} to decrypt the message and obtain m,

$$\mathcal{DEC}_{K_{PR}}(\mathcal{ENC}_{K_P}(m)) = m$$

One of the main properties of public key cryptosystems is that K_P and K_{PR} are mathematically related but it is infeasible to derive one from the other.

In the above explanation we have used the terms $\mathcal{ENC}$ and $\mathcal{DEC}$ as a general representation of the encryption and decryption functions that does not aim at being complete. Note that there are several public key schemes that can be used, and they all have their specific parameters and idiosyncrasies. Well-known public key cryptosystems are the RSA [12] and ElGamal [7]. Our proposal is independent of the cryptosystem and any of the mentioned above might be used.

[1]Note that the private key can be used to encrypt and the public key to decrypt. In this case instead of allowing private communications, the cryptographic system allows the signature (authentication) of messages.

III. Method

A. Rationale and desiderata

The use of information and communication technologies and, more specifically, mobile devices to monitor patients has significantly increased the chances of researchers to gather unprecedented amounts of timely data from multiple sources. However, in the context of m-health, if the data are not de-identified, patients have to consent that their data are released for secondary use. It has been shown that consent requirement can introduce bias in the analyses and their results [6]. Thus, it would be desirable to have architectures and methods allowing the automatic de-identification of distributed biomedical data to avert the need for patients' consent and the associated bias. There is a need for architectures to automatically gather biomedical data from multiple sources and to release them to third parties for secondary use while guaranteeing the privacy of patients.

Consequently, our goal is to design a simple and scalable architecture that uses an adapted microaggregation procedure and that allows the gathering of biomedical data from a number of distributed sources and their secure and private release to authorised parties. The following properties should be fulfilled by the architecture:

- **The architecture is not fully trusted.** The raw biomedical data sent by patients through the architecture should not be seen by any entity other than an authorized healthcare centre (or the like).
- **The resulting microdata sets should guarantee the privacy of patients** by assuring k-anonymity, where the security parameter k could be determined according to the specific needs of each microdata set. The disclosure risk should be kept low.
- **The released data set should be useful.** The information loss caused by the distributed microaggregation procedure should be low, and correlations between attributes should be maintained (as much as possible).

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

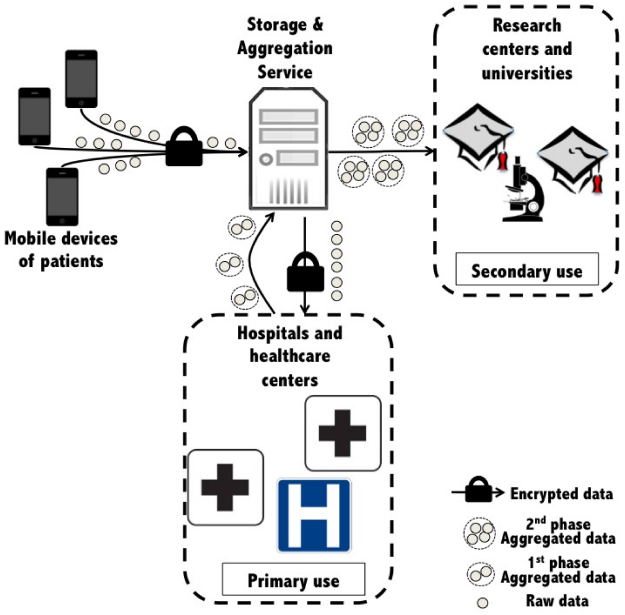


Fig. 1. Graphical scheme of the architecture and the double-phase microaggregation model. It can be observed that patients send data by means of mobile devices to a storage and aggregation service (SAS). Healthcare centres (HC) and research centres (RC) can access those data (raw or aggregated) depending on their intended use (primary or secondary).

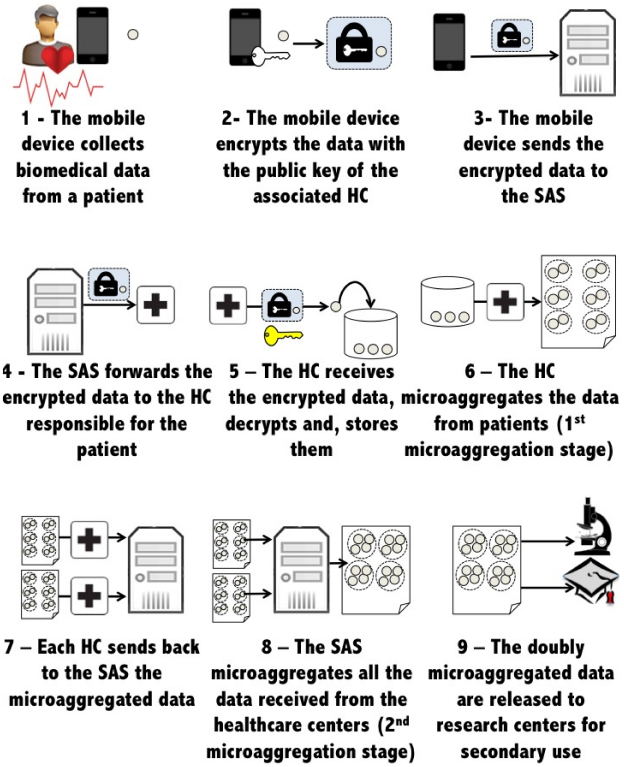


Fig. 2. Graphical description of the protocol and the flow of data within the proposed architecture.

B. Proposed architecture and assumptions

The proposed architecture considers four main actors, namely mobile devices (patients), healthcare centres (HC), research centres and a centralized storage and aggregation server (SAS) (see Figure 1 for a graphical representation of the architecture). These actors/entities interact so as to guarantee the private collection and sharing of data. We assume the following:

- Patients have mobile devices with communication capabilities able to collect data and encrypt them by using a public key cryptosystem (e.g. RSA [12]).
- The data collected from patients have the same attributes for all patients participating in a given study/trial. Note that this is quite common because clinical studies tend to be highly parametrised and strict with the treated variables.
- Each healthcare centre (doctors) has a pair of keys $\{K_P, K_{PR}\}$ of a public key cryptosystem (e.g. RSA [12]).
- Each patient is assigned to a healthcare centre that is responsible for the patient and his/her data. Although the interaction of patients with healthcare centres is kept to the minimum to maximise the benefits of using ICT, in most studies, regular supervision is required.
- Patients know the public key (K_P) of the healthcare centre to which they are assigned, and they share the same cryptosystem.

C. Protocol and dataflow

Figure 2 graphically depicts the different actions carried out by each actor of the system to gather and share data privately. First, mobile devices, carried by patients, collect biomedical data (and possibly other kinds of data) from patients (See step 1 of Figure 2). We refer to these data as,

$$\mathcal{D} = (d_{u_1}, d_{u_2}, \cdots, d_{u_i}, \cdots, d_{u_n})$$

where d_{u_i} represents the data of user/patient u_i. After collecting the data, the mobile device of each patient encrypts the gathered data with the public key ($K_P^{HC_i}$) of the healthcare center (HC_i) to which he/she is assigned (See step 2 of Figure 2) and generates a message (m) with the following information:

- User ID: u_i
- Healthcare Centre ID: HC_i
- Encrypted data: $\mathcal{ENC}_{K_P^{HC_i}}(d_{u_i})$

After encrypting the data and generating the message described above, the mobile device sends m to the storage and aggregation server (SAS) (See step 3 of Figure 2). Once the SAS receives the patients' data, it checks the healthcare centre ID of each message and forwards it to right HC_i (See step 4 of Figure 2). Note that the SAS cannot access to the raw data of the patients because it is encrypted with the public key of HC_i. Thanks to the use of their private key, each healthcare centre can decrypt and access the raw biomedical data of the patients for which they are responsible (See step 5 of Figure 2):

$$d_{u_i} = \mathcal{DEC}_{K_{PR}^{HC_i}}\left(\mathcal{ENC}_{K_P^{HC_i}}(d_{u_i})\right)$$

By doing so, doctors can analyse the complete data and decide on the proper procedures and protocols to apply without

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

any information loss (in this regard we assume that healthcare centres are trusted). After receiving and decrypting the data from the patients participating in a given trial/study, each healthcare center microaggregates all the data (by using a microaggregation algorithm such as MDAV or V-MDAV) with a given security parameter $HC(k)$ (See step 6 of Figure 2) and sends the resulting microaggregated data set back to the SAS (See step 7 of Figure 2).

When the SAS receives the microaggregated data sets from all healthcare centres, it merges them all and microaggregates them again by using again a microaggregation algorithm such as MDAV or V-MDAV, with a given security parameter $SAS(k)$ (See step 8 of Figure 2).

Thanks to this microaggregation procedure the finally obtained sanitized data set fulfils the k-anonymity property with $k = SAS(k)$, and it can be accessed by research centres for secondary use without requiring an explicit consent by the patients (See step 9 of Figure 2).

IV. Results and Discussion

The proposed architecture guarantees the private exchange of data between patients and healthcare centres through the SAS thanks to the use of public key cryptography. In this regard, even if the SAS is not a trusted party the architecture remains private. However, to protect the privacy of the patients with respect to research centres, public key cryptography is not suitable and we propose the use of our double-phase microaggregation procedure.

With the aim to assess the usefulness of the proposed double-phase microaggregation we have performed the following experiments:

- Compare the performance of a fixed-size microaggregation algorithm (MDAV) [5], a variable-size microaggregation algorithm (V-MDAV) [15], and a hybrid microaggregation algorithm based on genetic algorithms and MDAV (GA) [16] over a biomedical data set.
- Compare the performance in terms of information loss and disclosure risk of single-phase (classical) microaggregation and double-phase microaggregation based on MDAV and V-MDAV.
- Compare the ability of preserving correlations of single-phase and double-phase microaggregation.

To assess the performance of those methods we use a standard approach to compute the information loss (IL) as the ratio between the Sum of Square Errors (SSE) and the Sum of Square errors Total (SST). The SSE is computed as:

$$SSE = \sum_{i=1}^{s} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i) \quad (1)$$

where s is the number of groups in the k-partition, n_i is the number of elements in the i-th group, $\mathbf{x}_{ij}$ is the j-th element in the i-th group and $\bar{\mathbf{x}}_i$ is the centroid of the i-th group. SST is computed as:

$$SST = \sum_{i=1}^{n} (\mathbf{x}_i - \bar{x})'(x_i - \bar{x}) \quad (2)$$

where n is the number of records in the data set and $\bar{x}$ is the centroid of the data set.

Also, we evaluate the disclosure risk (DR) with a standard measure based on record linkage [11]. Considering that an attacker has a given subset of quasi-identifiers, the probability of re-identification success is measured. For our data set (described below), we have computed the average probability (expressed in %) considering that the attacker has different combinations of quasi-identifier (Age, Height, Weight, Age-Height, Age-Weight, Height-Weight and Age-Height-Weight).

A. Data set

Due to privacy issues, we have not used biomedical data belonging to real patients. However, with the aim to provide consistent evidence of the usefulness of our proposal, we have created a synthetic but realistic biomedical data set, based on real data. Our data set consists of three quasi-identifiers (i.e. age, weight and height) and five outcome attributes: maximum blood pressure (systolic), minimum blood pressure (diastolic), heartbeat rate, blood glucose level and blood oxygen saturation. A sample of three records of our data set is shown in Table II. Note that no identifiers (e.g. SSN or full names) are included in the data set.

TABLE II
Example of three records from our synthetic data set of biomedical data gathered through portable devices

Age	Height (cm)	Weight (kg)	Systolic Pres. (mmHg)	Diastolic Pres. (mmHg)	Heart-beat Rate	Glucose	Oxygen Satu-ration (%)
44	178.3	82.6	132	87	69	85	98
86	162.8	69.9	141	92	66	89	97
31	154.5	104.5	158	103	78	105	96

- **Age, weight and height:** These quasi-identifiers are real data sampled from The National Health and Nutrition Examination Survey III (NHANES-III [2]) like in [2]. More specifically, we have sampled 5,000 records.
- **Systolic and diastolic blood pressure:** Systolic pressure (SP) has been obtained by sampling a Gaussian distribution, whose average and standard deviation vary depending on the age of the subjects [3]. Diastolic pressure (DP) is computed as a function of the systolic pressure. In our case, DP=16.83 + 0.52(SP). Some random noise with null average and standard deviation equalling one is added to DP.
- **Heartbeat rate:** It has been generated independently of the age by considering a Gaussian distribution with average value equalling 72 and a standard deviation of 7.5. In this way, 95% of the samples are comprised between 57 and 87 [4].

[2]The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. More information can be found in http://www.cdc.gov/nchs/nhanes.htm

[3]These values are taken from http://www.highbloodpressureinfo.org/normal-blood-pressure-range.html

[4]Based on the data found in http://www.livestrong.com/article/134368-what-is-normal-resting-heart-rate/

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

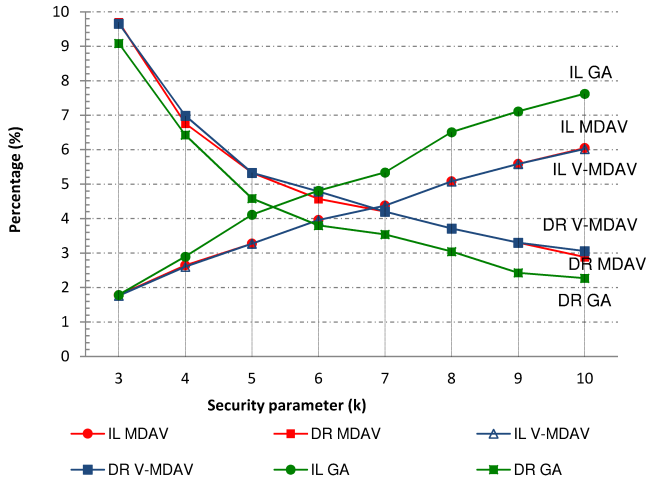


Fig. 3. Information loss (IL) and disclosure risk (DR) for MDAV, V-MDAV and GA for several values of k.

- **Blood glucose level:** In general, blood glucose levels are considered to be normal in the range between 70 and 100 mg/dL. Hence, we have generated those levels by sampling a Gaussian distribution having an average value of 85 and a standard deviation of 7.5.
- **Oxygen saturation:** Normal values of oxygen saturation are above 95%. Thus, we have generated those values by sampling a Gaussian distribution having an average value of 97 and a standard deviation of 1.25. These values are truncated (if necessary) so as not to exceed 100%.

B. MDAV vs V-MDAV vs. GA

Our first experiment consists in analysing the performance of MDAV, V-MDAV and GA over a sample of 1000 records of our biomedical data set. The aim of this comparison is to determine which of these methods is the best candidate to be implemented later as the main component of the double-phase microaggregation algorithm of our architecture.

As it can be observed in Figure 3 the IL and DR evolve inversely. The lower the IL the higher the DR. Also, it is clear that MDAV and V-MDAV perform very similarly and better than the GA in terms of IL with a reasonable DR. Thus, we choose the MDAV and the V-MDAV for our double-phase microaggregation. Due to space limitations, the complete set of results can be found in the Appendix A.

C. Classical vs. double-phase microaggregation

When classical microaggregation is used a single security parameter (k) has to be chosen[5]. However, with our double-phase microaggregation healthcare centres choose a security parameter $HC(k)$ and the SAS chooses another one $SAS(k)$. The final microaggregated data set fulfils the k-anonymity property where $k = SAS(k)$. We assume that $HC(k) \leq SAS(k)$.

[5]The value of k is determined by a security/privacy expert depending on the needs of each data set. In general the values of k lie between 3 and 10.

The first microaggregation phase is aimed at hiding the raw data from the SAS, thus $k = 2$ and $k = 3$ are considered to be enough. On the other hand, the second microaggregation phase leads to the final data set that is going to be released, in this case we have considered that $k = 4$ and $k = 6$ would be enough. Although other values are possible those are realistic choices and allow us to show the usefulness of our proposal.

TABLE III
PERCENTAGE OF IL AND DR FOR SEVERAL VALUES OF K AND FOR A DIFFERENT NUMBER OF PARTICIPANTS (HC). THE REFERENCE BASE LINE (WITH CLASSICAL SINGLE-PHASE MICROAGGREGATION) IS HIGHLIGHTED IN YELLOW

Method	HC(k) SAS(k)	HC + SAS	IL(%)	DR(%)
MDAV	2 - 4	4 + 1	1.3862	4.3114
	2 - 4	3 + 1	1.256	4.5086
	2 - 4	2 + 1	1.0648	5.2943
	4	-	0.9307	5.7943
	3 - 6	4 + 1	2.3099	2.6086
	3 - 6	3 + 1	2.023	3.0171
	3 - 6	2 + 1	1.7006	3.4914
	6	-	1.4132	3.8086
V-MDAV	2 - 4	4 + 1	1.3862	4.3114
	2 - 4	3 + 1	1.2557	4.5143
	2 - 4	2 + 1	1.0648	5.2943
	4	-	0.9272	5.7057
	3 - 6	4 + 1	2.3052	2.6229
	3 - 6	3 + 1	2.0223	3.0143
	3 - 6	2 + 1	1.7001	3.4886
	6	-	1.3943	3.8286

The results of the experiment on our complete data set with 5,000 records are shown in Table III. We have considered the situation in which the number of HC varies between 2, 3 and 4. It can be observed that the IL grows with the number of participants. However, the increase with respect to single-phase microaggregation is minimal (less than 1% in all tests).

D. Analysis of correlations preservation

The analysis of correlations between all the attributes in the data set reveals that they are all preserved and the use of double-phase microaggregation does not affect the correlations present in the original data set more than classical microaggregation. In fact most absolute differences between the correlations present in the original data set and the correlations in the distorted data sets are very close to zero in all cases. To emphasise this point Table IV shows the maximum absolute differences between correlations depending on the method used. It can be observed that the differences are very small always below 0.01, and that double-phase microaggregation does not affect the correlations.

The complete set of results of the correlation analysis can be found in Appendix B.

V. CONCLUSION

We have proposed an architecture that allows the private gathering and sharing of biomedical data in the context of

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TABLE IV
MAXIMUM ABSOLUTE DIFFERENCES BETWEEN THE CORRELATIONS IN THE ORIGINAL DATA SET AND THE CORRELATIONS IN THE DISTORTED DATA SETS FOR THE PROPOSED METHODS

Method	Max. Absolute Difference
MDAV-Single k=4	0.005166882
V-MDAV-Single k=4	0.009587946
MDAV-Double HC=2, HC(k)=2 SAS(k=4)	0.004323258
V-MDAV-Double HC=2, HC(k)=2 SAS(k=4)	0.006435297
MDAV-Double HC=3, HC(k)=2 SAS(k=4)	0.003748311
V-MDAV-Double HC=3, HC(k)=2 SAS(k=4)	0.003766998
MDAV-Double HC=4, HC(k)=2 SAS(k=4)	0.003602578
V-MDAV-Double HC=4, HC(k)=2 SAS(k=4)	0.003545481
MDAV-Single k=6	0.004303061
V-MDAV-Single k=6	0.006577712
MDAV-Double HC=2, HC(k)=3 SAS(k=6)	0.005018052
V-MDAV-Double HC=2, HC(k)=3 SAS(k=6)	0.005869891
MDAV-Double HC=3, HC(k)=3 SAS(k=6)	0.004015088
V-MDAV-Double HC=3, HC(k)=3 SAS(k=6)	0.003879271
MDAV-Double HC=4, HC(k)=3 SAS(k=6)	0.005114913
V-MDAV-Double HC=4, HC(k)=3 SAS(k=6)	0.004991748

m-health. We have introduced to concept of double-phase microaggregation to limit the information accessible by intermediate entities (such as the SAS). Our experiments show that MDAV and V-MDAV perform similarly over our biomedical data set. In addition, we have observed that the effect of double-phase microaggregation over the IL is negligible with respect to the classical microaggregation and that it preserves the correlations of the original data set. Then, we can conclude that the distributed double-phase microaggregation proposed can be applied in a distributed environment to protect the privacy of individuals with the same effects of classical microaggregation.

Although we have applied our proposal over biomedical data, the architecture and method proposed might be smoothly applied to other fields like economics, tourism, energy, and so on.

Further research might include the analysis of the influence of time in the series of data collected using our model. It is possible that the use of time information might help attackers to get extra knowledge and increase the probability of re-identification. Also, the use and protection of location data should be considered in the future.

ACKNOWLEDGEMENTS

This work was partly funded by the Spanish Government through project CONSOLIDER INGENIO 2010 CSD2007-0004 “ARES”, project TSI2007-65406-C03-01 “E-AEGIS”, project TIN2011-27076-C03-01 “CO-PRIVACY”, and by the Government of Catalonia under grant 2009SGR1135. The authors are with the UNESCO Chair in Data Privacy, but the views expressed in this article are their own and do not commit UNESCO.

REFERENCES

[1] Brand R. Microdata protection through noise addition. *Lecture Notes in Computer Science*. 2002;2316:97-116.

[2] Brown C.D., Higgins M., Donato K.A., et al. Body mass index and prevalence of hypertension and dyslipidemia. *Obesity Research*. 2000;8(9):605-619.

[3] Dalenius T., Reiss S.P. Data-swapping: A Technique for Disclosure Control. *Journal of Statistical Planning and Inference*. 1982;6(1):73-85.

[4] Domingo-Ferrer J., Sebé F., Castellà J. On the security of noise addition for privacy in statistical databases. *Lecture Notes in Computer Science*. 2004;3050:149-161.

[5] Domingo-Ferrer J., Martínez-Ballesté A., Mateo-Sanz J.M., et al. Efficient Multivatiate Data-Oriented Microaggregation. *Int. J Very Large Databases* 2006;15(4):355-369.

[6] El Emam K., Kalman Dankar F., Issa R., et al. Globally Optimal k-Anonymity for De-Identification of Health Data. *J Am Med Inform Assoc* 2009;16(5):670-82.

[7] ElGamal T., A Public-key Cryptosystem and a Signature Scheme Based on Discrete Logarithms. *IEEE Transactions on Information Theory*.1985.31(4):469-472.

[8] Greenberg B. Rank swapping for masking ordinal microdata. *Tech. report. U.S. Bureau of the Census* (unpublished manuscript). 1987.

[9] Naehrig M., Lauter K., Vaikuntanathan V. Can homomorphic encryption be practical?. In Proceedings of the 3rd ACM workshop on Cloud computing security workshop (CCSW ’11). ACM, New York, NY, USA, 113-124.

[10] Matthews G.J., Harel O., Data confidentiality: A review of methods for statistical disclosure limitation and methods for assessing privacy. *Statistics Surveys*. 2011;5:1?29.

[11] Pagliuca D., Seri G., Some Results of Individual Ranking Method on the System of Enterprise Accounts Annual Survey. *Esprit SDC Project*, Deliverable MI3/D2. 1998.

[12] Rivest R., Shamir A., Adleman L., A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM*.1978;21(2):120-126.

[13] Samarati P., Sweeney L. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. *Proceedings of the IEEE Symposium on Research in Security and Privacy* (S&P). 1998.

[14] Samarati P. Protecting Respondents’ Identities in Microdata Release. *IEEE Transactions on Knowledge and Data Engineering*. 2001;13(6):1010-1027.

[15] Solanas A., Martínez-Ballesté A., V-MDAV: Variable group size multivariate microaggregation. *COMPSTAT’2006*: 917-925.

[16] Solanas A., Martínez-Ballesté A., González-Nicolas Ú., A variable-MDAV-based partitioning strategy to continuous multivariate microaggregation with genetic algorithms. *International Joint Conference on Neural Networks(IJCNN)*.2010:1-7.

[17] Sweeney L. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*. 2002;10(5):557-570.

[18] Willenborg L., DeWaal T. Statistical Disclosure Control in Practice. *Lecture Notes in Statistics*, 111. Springer-Verlag. 1996.

[19] Willenborg L., DeWaal T. Elements of Statistical Disclosure Control. *Lecture Notes in Statistics*, 155. Springer-Verlag. 2001.

APPENDIX A
COMPLETE RESULTS SET FOR MDAV, V-MDAV AND GA

The complete results are shown in Table V. The notation used is the following:

- k is the security parameter.
- IL is the information loss (in %).
- A is the probability (in %) of re-identification based on record linkage when the attacker knows the quasi-identifier (AGE).
- H is the probability (in %) of re-identification based on record linkage when the attacker knows the quasi-identifier (HEIGHT).
- W is the probability (in %) of re-identification based on record linkage when the attacker knows the quasi-identifier (WEIGHT).

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

- $A-H$ is the probability (in %) of re-identification based on record linkage when the attacker knows the pair of quasi-identifiers (AGE and HEIGHT).
- $A-W$ is the probability (in %) of re-identification based on record linkage when the attacker knows the pair of quasi-identifiers (AGE and WEIGHT).
- $H-W$ is the probability (in %) of re-identification based on record linkage when the attacker knows the pair of quasi-identifiers (HEIGHT and HEIGHT).
- $A-H-W$ is the probability (in %) of re-identification based on record linkage when the attacker knows the all the quasi-identifiers. (AGE, HEIGHT and WEIGHT).
- DR is the disclosure risk (in %), computed as the average of (A, H, W, A-H, A-W, H-W and A-H-W).

TABLE V
COMPLETE RESULTS FOR THE MDAV VS V-MDAV VS. GA SECTION.

Meth.	k	IL	A	H	W	A H	A W	H W	A H W	DR
MDAV	3	1.77	1.3	2	1.4	10.3	10.1	10.8	32	9.7
	4	2.64	1	1.6	1.3	6.1	6.8	7.3	23.2	6.76
	5	3.27	1	0.7	0.8	4.3	6.3	5	19.2	5.33
	6	3.96	0.1	0.9	1.4	3.6	5.1	4.9	16	4.57
	7	4.37	0.5	1.1	1	4.4	3.5	4.8	14.1	4.2
	8	5.08	0.4	0.9	0.9	3.9	4.2	3.5	12.2	3.71
	9	5.59	0.8	0.6	1.2	4	2.7	3	10.8	3.3
	10	6.05	0.1	1	0.4	3	2.4	3.3	10	2.89
VMDAV	3	1.76	1	2	1.6	10.5	10.8	10.9	30.8	9.66
	4	2.60	0.6	1.3	1.2	6.9	8.1	7.5	23.3	6.99
	5	3.27	1	0.7	0.8	4.3	6.3	5	19.2	5.33
	6	3.96	0.6	1.2	1	4.5	5.2	5.2	15.8	4.79
	7	4.37	0.5	1.1	1	4.4	3.5	4.8	14.1	4.2
	8	5.08	0.4	0.9	0.9	3.9	4.2	3.5	12.2	3.71
	9	5.59	0.8	0.6	1.2	4	2.7	3	10.8	3.3
	10	6.02	0.5	0.9	0.4	3.1	3	3.8	9.7	3.06
GA	3	1.79	1	1.8	1.3	9.5	10.1	10.7	29.2	9.09
	4	2.90	1.3	1.5	1.4	6.8	7.3	6.9	19.8	6.43
	5	4.12	0.9	0.9	1	5.2	4.6	4.6	14.9	4.59
	6	4.81	0.4	0.8	1.1	3.1	4.3	4.7	12.2	3.8
	7	5.34	0.5	0.9	0.9	3.4	4.1	4	11	3.54
	8	6.51	0.4	0.9	0.8	3.1	3.2	3.5	9.4	3.04
	9	7.11	0.7	0.5	1.1	2	2	2.9	7.8	2.43
	10	7.62	0.3	0.4	0.7	2.4	2.2	2.9	7	2.27

APPENDIX B
CORRELATIONS ANALYSIS BETWEEN THE QUASI-IDENTIFIERS AND ALL THE ATTRIBUTES FOR THE PROPOSED METHODS

Tables VI and VII show the results of the analysis of correlations.



Agusti Solanas is a researcher at the CRISES Research Group and the UNESCO Chair in Data Privacy in the Department of Computer Science and Mathematics at the Rovira i Virgili University (URV) of Tarragona, Catalonia, Spain. He received his B.Sc. and M.Sc. degrees in Computer Engineering from URV in 2002 and 2004, respectively,the latter with honours (Outstanding Graduation Award). He received a Diploma of Advanced Studies (Master) in Telematics Engineering from the Technical University of Catalonia (UPC) in 2006. He received a Ph.D. in Telematics Engineering from the Technical University of Catalonia in 2007 with honours (A cum laude). His fields of activity are privacy, security, clustering, neural networks and evolutionary computation. He has participated in several European-, Spanish- and Catalan-funded research projects. He has authored over 70 publications and he has delivered several invited talks. He has served as Chair, programme committee member and reviewer in several conferences and journals. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Association for Computing Machinery (ACM).



Antoni Martínez-Ballesé received the M.Sc. in Computer Science and Engineering from Universitat Rovira i Virgili, (Catalonia, Spain, 2001), and Ph.D. degree in Telematics from Universitat Politcnica de Catalunya, (Catalonia, Spain, 2004). He is a tenured assistant professor at the Universitat Rovira i Virgili. His research interests are security and privacy in ICT and specially the privacy of their users. Since 2001, he is with CRISES research group where he has taken part in several research projects. Moreover, he is serving as vice-dean for Computer Science and Engineering and collaborates with the Universitat Oberta de Catalunya.



Josep M. Mateo-Sanz is a tenured assistant professor at Rovira i Virgili University. He got the M.Sc. and the Ph.D. in mathematics at University of Barcelona in 1992 and 1998, respectively. His fields of activity are data privacy and data security using statistical techniques. In 2010, he received the Teaching Quality award of the Rovira i Virgili University. In the same year, he received the Jaume Vicens Vives award of the Generalitat de Catalunya to the University Teaching Quality. He has published over 40 works, one of them was ISI Highly Cited Paper at the beginning of 2005. He has also authored more than 50 congress contributions. He has participated in various Spanish, European and North American projects. He has also participated in research contracts with the U.S. Bureau of the Census, Yahoo Inc. and the Statistical Institute of Catalonia, among others. He is currently a member of the CONSOLIDER ARES Group.

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TABLE VI
CORRELATIONS ANALYSIS BETWEEN THE QUASI-IDENTIFIERS AND ALL THE ATTRIBUTES FOR THE PROPOSED METHODS (1/2)

ORIGINAL DATA SET	**Age**	**Weight**	**Height**
Age	1	-0.114371504	-0.177608513
Weight	-0.114371504	1	0.470844055
Height	-0.177608513	0.470844055	1
Systolic BP	0.73726533	0.12961159	-0.171904118
Diastolic BP	0.730060692	0.097707698	-0.166204079
Heartbeat rate	-0.013536149	0.006403094	-0.000681947
Blood Glucose	0.009678634	0.005741966	0.020522736
Oxygen saturation	-0.006147947	-0.004696471	0.001579498

MDAV-Single k=4	**Age**	**Weight**	**Height**
Age	1	-0.11523395	-0.179477693
Weight	-0.11523395	1	0.476010937
Height	-0.179477693	0.476010937	1
Systolic BP	0.735167393	0.127611633	-0.171975727
Diastolic BP	0.727474356	0.0963951	-0.16622132
Heartbeat rate	-0.013147262	0.008241806	-0.000505785
Blood Glucose	0.009905156	0.005789089	0.021207719
Oxygen saturation	-0.005434974	-0.00526146	-0.000375406

V-MDAV-Single k=4	**Age**	**Weight**	**Height**
Age	1	-0.115181442	-0.17957088
Weight	-0.115181442	1	0.475920575
Height	-0.17957088	0.475920575	1
Systolic BP	0.736035957	0.128669735	-0.171322347
Diastolic BP	0.728509734	0.09726474	-0.165588334
Heartbeat rate	-0.012771095	0.007246755	-0.001242341
Blood Glucose	0.011275013	0.005634182	0.020221065
Oxygen saturation	-0.004448055	-0.008081708	0.000288412

MDAV-Double HC=2, HC(k)=2 SAS(k=4)	**Age**	**Weight**	**Height**
Age	1	-0.114428841	-0.18066854
Weight	-0.114428841	1	0.478729967
Height	-0.18066854	0.478729967	1
Systolic BP	0.73560557	0.125354055	-0.169674385
Diastolic BP	0.728660496	0.09430094	-0.163643146
Heartbeat rate	-0.013311167	0.005920054	-0.002370725
Blood Glucose	0.008737942	0.007412701	0.01688446
Oxygen saturation	-0.007362043	-0.001646411	0.002083137

V-MDAV-Double HC=2, HC(k)=2 SAS(k=4)	**Age**	**Weight**	**Height**
Age	1	-0.114428841	-0.18066854
Weight	-0.114428841	1	0.478729967
Height	-0.18066854	0.478729967	1
Systolic BP	0.73560557	0.125354055	-0.169674385
Diastolic BP	0.728660496	0.09430094	-0.163643146
Heartbeat rate	-0.013311167	0.005920054	-0.002370725
Blood Glucose	0.008737942	0.007412701	0.01688446
Oxygen saturation	-0.007362043	-0.001646411	0.002083137

MDAV-Double HC=3, HC(k)=2 SAS(k=4)	**Age**	**Weight**	**Height**
Age	1	-0.114105318	-0.182102678
Weight	-0.114105318	1	0.479588991
Height	-0.182102678	0.479588991	1
Systolic BP	0.733560133	0.125752527	-0.170173762
Diastolic BP	0.726386086	0.094393402	-0.164303825
Heartbeat rate	-0.01057238	0.007087723	9.49E-05
Blood Glucose	0.008404068	0.007228893	0.020632771
Oxygen saturation	-0.005881683	-0.002988496	-0.001573881

V-MDAV-Double HC=3, HC(k)=2 SAS(k=4)	**Age**	**Weight**	**Height**
Age	1	-0.114112778	-0.182089698
Weight	-0.114112778	1	0.479593005
Height	-0.182089698	0.479593005	1
Systolic BP	0.733516167	0.125713681	-0.170210349
Diastolic BP	0.726353216	0.094350811	-0.164331164
Heartbeat rate	-0.010525694	0.007128748	1.52E-05
Blood Glucose	0.008475467	0.007216736	0.020651459
Oxygen saturation	-0.005788299	-0.003036082	-0.001516784

MDAV-Double HC=4, HC(k)=2 SAS(k=4)	**Age**	**Weight**	**Height**
Age	1	-0.114934421	-0.182290711
Weight	-0.114934421	1	0.480124307
Height	-0.182290711	0.480124307	1
Systolic BP	0.735060351	0.12614356	-0.17304666
Diastolic BP	0.728122126	0.095313602	-0.167201332
Heartbeat rate	-0.011624007	0.006872884	0.001338018
Blood Glucose	0.008057121	0.005394871	0.018420264
Oxygen saturation	-0.005349106	-0.003442382	0.002028697

V-MDAV-Double HC=4, HC(k)=2 SAS(k=4)	**Age**	**Weight**	**Height**
Age	1	-0.114934421	-0.182290711
Weight	-0.114934421	1	0.480124307
Height	-0.182290711	0.480124307	1
Systolic BP	0.735060351	0.12614356	-0.17304666
Diastolic BP	0.728122126	0.095313602	-0.167201332
Heartbeat rate	-0.011624007	0.006872884	0.001338018
Blood Glucose	0.008057121	0.005394871	0.018420264
Oxygen saturation	-0.005349106	-0.003442382	0.002028697

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TABLE VII
CORRELATIONS ANALYSIS BETWEEN THE QUASI-IDENTIFIERS AND ALL THE ATTRIBUTES FOR THE PROPOSED METHODS (2/2)

MDAV-Single k=6	**Age**	**Weight**	**Height**	**V-MDAV-Single k=6**	**Age**	**Weight**	**Height**
Age	1	-0.116133162	-0.179995689	**Age**	1	-0.116204896	-0.180310575
Weight	-0.116133162	1	0.478254649	**Weight**	-0.116204896	1	0.477955649
Height	-0.179995689	0.478254649	1	**Height**	-0.180310575	0.477955649	1
Systolic BP	0.733118287	0.12675717	-0.17004796	**Systolic BP**	0.733959899	0.125952041	-0.170851072
Diastolic BP	0.725877126	0.09548688	-0.164712568	**Diastolic BP**	0.726690391	0.094793702	-0.166185741
Heartbeat rate	-0.015927069	0.003824026	-0.001733812	**Heartbeat rate**	-0.015880625	0.002970123	-0.005239694
Blood Glucose	0.010784467	0.007351025	0.020723089	**Blood Glucose**	0.010365171	0.007051112	0.021417581
Oxygen saturation	-0.003271318	-0.006800687	0.000250379	**Oxygen saturation**	-0.005301375	-0.005667807	-0.000588808
MDAV-Double HC=2, HC(k)=3 SAS(k=6)	**Age**	**Weight**	**Height**	**V-MDAV-Double HC=2, HC(k)=3 SAS(k=6)**	**Age**	**Weight**	**Height**
Age	1	-0.115716609	-0.182164493	**Age**	1	-0.115706164	-0.182157784
Weight	-0.115716609	1	0.48205058	**Weight**	-0.115706164	1	0.482055281
Height	-0.182164493	0.48205058	1	**Height**	-0.182157784	0.482055281	1
Systolic BP	0.732296124	0.123470214	-0.170928273	**Systolic BP**	0.732282036	0.123454193	-0.170917446
Diastolic BP	0.725153508	0.09227121	-0.165440665	**Diastolic BP**	0.725140549	0.092242503	-0.165435196
Heartbeat rate	-0.013458105	0.003715133	0.000559023	**Heartbeat rate**	-0.013401572	0.003783178	0.000630198
Blood Glucose	0.007861142	0.004467267	0.015705037	**Blood Glucose**	0.00789841	0.004640561	0.015807692
Oxygen saturation	-0.004825102	-0.004161671	-0.00050589	**Oxygen saturation**	-0.004794742	-0.004149835	-0.00047096
MDAV-Double HC=3, HC(k)=3 SAS(k=6)	**Age**	**Weight**	**Height**	**V-MDAV-Double HC=3, HC(k)=3 SAS(k=6)**	**Age**	**Weight**	**Height**
Age	1	-0.114477067	-0.182218878	**Age**	1	-0.114447519	-0.182222832
Weight	-0.114477067	1	0.483724198	**Weight**	-0.114447519	1	0.483707482
Height	-0.182218878	0.483724198	1	**Height**	-0.182222832	0.483707482	1
Systolic BP	0.733618838	0.123303795	-0.168909815	**Systolic BP**	0.733577298	0.123256336	-0.168866789
Diastolic BP	0.727040366	0.091798578	-0.163598362	**Diastolic BP**	0.72698834	0.091734573	-0.163510163
Heartbeat rate	-0.013280477	0.00773022	0.001215116	**Heartbeat rate**	-0.013324891	0.007662449	0.001188028
Blood Glucose	0.007532137	0.004758595	0.019030043	**Blood Glucose**	0.007542667	0.004789036	0.019288162
Oxygen saturation	-0.006026761	-0.003498591	0.000122293	**Oxygen saturation**	-0.006031049	-0.00333856	-9.79E-05
MDAV-Double HC=4, HC(k)=3 SAS(k=6)	**Age**	**Weight**	**Height**	**V-MDAV-Double HC=4, HC(k)=3 SAS(k=6)**	**Age**	**Weight**	**Height**
Age	1	-0.115620665	-0.183165619	**Age**	1	-0.115635058	-0.183170574
Weight	-0.115620665	1	0.485508521	**Weight**	-0.115635058	1	0.485499534
Height	-0.183165619	0.485508521	1	**Height**	-0.183170574	0.485499534	1
Systolic BP	0.730648345	0.125738053	-0.16961598	**Systolic BP**	0.730771166	0.125771653	-0.169435165
Diastolic BP	0.724056642	0.094486114	-0.163795885	**Diastolic BP**	0.72412441	0.094530117	-0.163590492
Heartbeat rate	-0.013088454	0.002920223	0.000143293	**Heartbeat rate**	-0.012962801	0.002859011	0.000199041
Blood Glucose	0.01264705	0.006633217	0.016440064	**Blood Glucose**	0.012534415	0.006592205	0.016403198
Oxygen saturation	-0.003994789	-0.004009294	0.001746623	**Oxygen saturation**	-0.004071356	-0.003673603	0.001837612

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.