Personalised medicine challenges: quality of data

Ricardo Cruz-Correia¹ · Duarte Ferreira¹ · Gustavo Bacelar¹ · Pedro Marques¹ · Priscila Maranhão¹

Received: 20 March 2017 / Accepted: 27 April 2018 / Published online: 2 June 2018
© The Author(s) 2018

Abstract

Personalised evidence-based-medicine aims to use stored health data to prevent future illnesses. This implies that data should be stored in a readable and understandable form, at least until the death of the person in question. The aim of this paper is to discuss the challenges that arise from the existing pressure to maintain health data in electronic format for many decades. Today clinical databases are filled with heterogeneous data regarding who has collected it, protocols used, detail, precision, and subjectivity. Some data elements are typically more exposed to these problems (e.g. diagnosis) than others (e.g. laboratory results). It is critical that data scientists fully understand how data were collected. Also, it is very important to store context information, protocols used and accuracy/precision information in clinical databases to ensure future understanding of such data.

Keywords Quality of data · Personalised medicine · Electronic patient record · Heterogeneous data

1 Introduction

Personalised evidence-based-medicine (EbM) uses stored health data, namely of patient diagnoses, laboratory work, insure claims, and demographic information among other. This information allows to move beyond the reactive approach of treating illness, allowing healthcare providers to predict and prevent future illnesses [14] and therefore become a promising application area for data science as a discipline. Nevertheless, this area has some specific challenges as the use

of such data may occur many years after or in a completely different setting than when and where it was collected.

Despite many efforts along the years for improving normalisation and standardization of clinical data, concerns regarding these aspects are still present in recent initiatives intending to push forward personalised medicine. Projects such as FP7 MyHealthAvatar [38] and DISCIPULUS [37] embody the relevance of having digital clinical information for pursuing personalised medicine thus reinforcing the importance on guarantying completeness regarding patient data allowing a complete view and integrated analysis of the patient health: to this end the methods used for the acquisition of information must be such that information is given as a standardised set of data and preferably provided with uncertainty ranges. This concerns continued to the next call within the Virtual Physiological Human (VPH) with more hands-on approach like (p-medicine: from data sharing and integration via VPH models to personalised medicine) [36]. Another relevant aspect, considered critical to the development of personalised medicine, is surfacing from recent initiatives like IGNITE (Implementing GeNomics In pracTicE) projects [5] where the priority for having genomic data as a part of the electronic health record is considered to be high. These projects need to deal with data quality and precision issues, data heterogeneity, and data aggregation to create a "big picture" representation of the patient.

NORTE-01-0145-FEDER-000016 (NanoSTIMA) is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

✉ Ricardo Cruz-Correia
rcorreia@med.up.pt

Duarte Ferreira
duarteng.ferreira@gmail.com

Gustavo Bacelar
gbacelar@gmail.com

Pedro Marques
pmarques@med.up.pt

Priscila Maranhão
priscilamaranhao@gmail.com

¹ CINTESIS - Center for Health Technology and Services Research, Rua Dr. Plácido da Costa, Porto, Portugal

Therefore when one considers health-related data, it becomes very important to consider the longevity of data regarding its usefulness and how will data age [7]. Due to the fact that we do not know for sure how data will be used in the future and therefore its value, the best way to protect such use opportunities is to store all data in a readable and understandable form, at least until the death of the person in question.

Take for example the data of newborns related to having low weight at childbirth. It is known that being small for gestational age (SGA) has a higher risk of short stature than children born at normal size do [40]. Besides, SGA has been associated with an increased prevalence of cardiovascular disease, essential hypertension, and metabolic disease, particularly type 2 diabetes mellitus [22]. In another example, the short duration of breastfeeding is significantly correlated with metabolic syndrome and obesity in childhood, being an important factor for preventing metabolic disorders [31,54]. In these cases, maintaining the data related to the birth weight and breastfeeding time of each newborn in their patient record may become very useful to calculate the risk of cardiovascular disease of each person many years after.

Considering the life expectancy at birth in 2014 in Europe is around 80.9 years, according to Eurostat [10], we should be aiming to use formats and structures that could sustain such a lifespan to store our own data. This means that data collected (e.g. the weight of newborns) in 2017 should be stored in a format understandable until 2098.

The aim of this paper is to discuss the difficulties and possible solutions to problems that rise from the existing pressure to maintain health data in electronic format for many decades.

2 Types of health data

Health data can be very different depending on its type (e.g. observations, evaluations, instructions), or if it is collected by humans (e.g. medical doctors, nurses, patients) or machines (e.g. ICU monitoring systems, digital thermometer). Some of these values are interpreted by humans before data input and therefore vulnerable to subjectivity, others are automatically sampled or aggregated (e.g. the mean blood pressure of the last 24 h). The protocol used to collect data is very rarely recorded, and in some settings, it may often change along the years (e.g. use of a different anatomical location, patient position, or device to measure the blood pressure).

OpenEHR is a standard (<http://www.openEHR.org>) that provides an open-source software infrastructure for implementing an electronic health record (EHR) in a clinical knowledge domain [53]. It is based on a multi-level, service-oriented architectural and follows a single-source modelling approach. At its core, there is a stable reference model that

defines the logical structures used in the modelling process. On top of the reference model structures, we have data points, which in openEHR are called archetypes, and are composable structures that define the way clinical data should be stored. Each archetype has an identifier, and each data point can be accessed through a path inside the archetype. These identifiers and paths are unique and independent of the context in which the archetype is used.

Archetypes can be combined in higher level, context-specific structures that are called templates. These templates can inherit all or only a part of each archetype's data points. This is a very powerful characteristic that allows re-usability in different contexts without losing the ability to compare data points between different templates that use the same archetype. Application developers can base persistence models, interfaces, and forms on openEHR templates.

In Table 1, different data types were grouped according to the types of clinical and administrative entries of openEHR standard. We also added a system usage data category that is not present in openEHR standard. This data category does not represent clinical or administrative data but data about the systems, errors, access logs, or how the system contact with other systems. Although not directly related to healthcare delivery, it is of the utmost importance for data provenance, traceability, and data security [8]. Thus, the purpose of each clinical entry is as follows [47]:

- *Observation* for recording information from the patient's world—anything measured by a clinician, a laboratory or by them, or reported by the patient as a symptom, event or concern;
- *Evaluation* for recording opinions and summary statements (usually clinical), such as problems, diagnoses, risk assessments, goals, etc., that are generally based on observation evidence;
- *Instruction* for recording orders, prescriptions, directives, and any other requested interventions;
- *Action* for recording actions, which may be due to Instructions, e.g. drug administrations, procedures, etc;
- *Administrative* for recording administrative events, e.g. admission, discharge, consent, etc.

Moreover, for each type of data represented in Table 1, a few examples are given to illustrate and a set of issues (with a – is a problem with + is an advantage) are itemised. Below follows a description of this issues.

- Observation data, e.g. temperature and blood pressure, when collected by:
 - *Humans* are subject to reading or interpretation errors when transcribing, subjectivity when value reading is

Table 1 Relation between types of data collected and issues affecting quality of data

Type of data	Examples	Collected by	
		Humans	Machines
Observations	Temperature	– Make transcriptions	+ Usually sensors
	Blood pressure	– Subjective	– Do not store protocol data
		– Lack of standards	
Evaluations	Diagnosis	– Subjective	
	Adverse reaction risk	– Protocols changes	
		– Need to teach terminology	
Instructions (orders)	Laboratory orders	+ Reliable	
Actions	Laboratory reports	– Incomplete data	+ Automatic analysis
	Surgery reports		
Administrative	Patient ID number, name	– Some basic data (eg. name) may change	+ Easier when based on ID cards
	Date of birth and gender		
System usage	Audit trails	– Security credential sharing	– Consistent time issues
	Messaging logs		

not clear [19], lack of standardization in the way they are stored.

- *Machines* the error level is known when using sensors, but there is still a lack of storage of data collection protocol.
- Evaluation data, e.g. diagnosis and adverse reaction risk, when collected by:
 - *Humans* are also subject to subjectivity [19], lack of standardization on the way terminology is used and understood, and the meaning of some disease concepts changes through time when we consider decades.
- Instruction data, e.g. laboratory orders, when collected by:
 - *Humans* these data are normally reliable in the sense that it describes what the person that filled the order really asked for. The use of terminologies, in this case, is well known and very common.
- Actions data, e.g. laboratory or surgery reports, when collected by:
 - *Humans* these data may suffer from incompleteness when introduced by humans, as it tends to be very verbose (many values) and we tend to value more the interpretation of those values than the values themselves.
 - *Machines* the automatic collection and analysis of action reports, when possible, have a great potential and is common in most healthcare settings.
- Administrative data, e.g. patient or visit identification and demographics data, when collected by:
 - *Humans* most systems depend critically on the correct identification of persons (both patients and health professionals). Incorrect identification is more common than most admit and can, obviously, have a huge impact on patient care. Furthermore, some data elements we use daily to identify persons univocally may not be that unique or change during our lifetime (eg. women's name change after marriage in many cultures).
 - *Machines* the widespread use of ID cards and other similar identification technologies have improved the quality of the data and also the time and effort to collect such data.
- System usage data, e.g. audit trails and messaging logs, when collected by:
 - *Humans* sharing credentials (e.g. login and password) or computer sessions are a major limitation to analyse audit trail information both for auditing or process mining purposes.
 - *Machines* even data that are collected automatically may have many problems. For instance, having consistent time in logs is still uncommon in many settings. Many servers do not have their time synchronised, and the logs do not use proper time/date standards (e.g. ISO8601) to deal with timezones or daylight saving time.

	PatID	Date	AP
+- Form in use since 1992 --+	254	03.05.92	0
[] Allergy to penicillin	255	04.05.92	1
+-----+	...	...	...
++- Form in use since 2002 -----+	1232	06.07.02	0
Allergy to penicillin: () Yes () No () Unknown	1233	07.07.02	1
+-----+	1234	08.07.02	9
	...	...	...
+- Form in use since 2013 -----+	8872	18.02.13	0
Allergy to penicillin:	8875	20.02.13	0
() Yes, confirmed by doctor () No, confirmed by doctor	8877	20.02.13	1
() Yes, informed by patient () No, informed by patient	8879	20.02.13	9
() Unknown	8901	21.02.13	2
+-----+	8903	22.02.13	3

Fig. 1 Illustration of how different versions of a form feed the same data table, making it difficult to interpret the meaning of each answer. AP means allergy to penicillin

3 Data collection form issues

As stated in [45], few problems are more challenging than the development of effective techniques for capturing patient data accurately, completely, and efficiently. Although more and more data are being collected using sensors or other automatic forms, most data existing in clinical databases are still the result of filling a form by a health professional or patient. These forms are present inside electronic patient records (EPRs) and typically have both very structured data entries and narrative entries to record patient data. The amount of structured data is dependent on the time it takes to fill it in, the importance of such data elements to the institution, and the difficulty to structure it in multiple data elements in opposition to leaving an open text field.

3.1 Form formats

One very important aspect related to the quality of data is the forms used to present and collect such data [52]. This has been realised in many clinical scenarios, and therefore, efforts must be made to standardise such forms.

There is evidence that question wording and framing, including the choice and order of response categories, can have an important impact on the nature and quality of responses [27]. McColl et al. also stated that through careful attention to the design and layout of questionnaires, the risk of errors in posing and interpreting questions and in recording and coding responses can be reduced, and potential inter-rater variability can be minimised.

Another example is a work aiming to define a synopsis format that is effective in delivering essential pathological information and to evaluate the aesthetic appeal and the impact of varying format styles on the speed and accuracy of data extraction [46]. One of its main conclusions is that

human factors engineering should be considered in the display of patient data.

3.2 Data values during form changes

To better illustrate the case, imagine that a particular hospital has been collecting data about allergy to penicillin of their patients (see Fig. 1). These data are being collected since 1992. During these years, the forms used to collect this data have changed, aiming to improve the quality of data. The software developers have chosen to have the allergy to penicillin values recorded in the same database field independently of the form used. Unfortunately, changing forms without changing the data structures where those values are recorded or storing the form version used to collect, is much more common than one would expect. In this (common) case, interpretation of such values will be much harder.

3.3 Paper versus electronic forms

One important difference between storing patient data in paper forms and computer systems is the fact that whilst in paper forms the questions and the answers are stored together (the paper form), in computer systems the questions only exist as software forms in an application and the answers are stored in the databases. Also, health institutions normally do not maintain the previous versions of the computer systems, easily leading to a situation where you have the answers provided by health professionals, but one does not know the exact questions that were made. Instead of the questions, one ends up with a list of field names that may not describe the question made to the user. Knowing the answers without knowing the exact questions is not useful and may be dangerous.

3.4 Data transformation

These issues (lack of formalism and clarity in data handling) produce a low rate of reproducibility in research [33]. The use of data provenance, which is a formal representation of computational processes, may be a solution for this issue. Complex computational tools to analyse with large quantities of data create the need for more precise descriptions of the origin of data the transformations that have been applied to those data, and the implications of the results. Pasquier et al. suggestion of publishing the source code used for data transformations in scientific papers, could be extended to also include the source code of the systems used in data collection.

Data provenance refers to attributes of the origin of information, it can help in guaranteeing proof of data integrity [17], which is very useful when using cloud environments or storing data for long periods of time.

4 Storing the error level

Databases are used to store facts. The data should have enough precision, detail, and context to be properly understood and analysed. In the healthcare domain, data can be collected using many different protocols and devices through time.

4.1 Medical devices

A large portion of medical data is originated in medical devices. Like all other calibrated devices, they have the capability of measuring physical properties just to a certain accuracy and precision. Obviously, no measurement is perfect and all have some error associated with them. Accuracy and precision are necessary to ensure that results are valid.

As an example, there are few studies addressing the reliability of home blood pressure monitoring devices and the quality of its data. Jung et al. present a study that aimed to evaluate the current status of home BP devices in terms of validation and accuracy [21]. This study showed that non-validated devices are used widely in clinical practice and a substantial portion is inaccurate. Storing the capability of the device to measure the reality is also important to properly use patient data in the future.

4.2 Medical records

This reality can also be found when collecting more subjective information in medical records. In these cases, knowing the exact source of the data, e.g. doctor opinion or patient description, may have an impact on the interpretation of data. It should be considered the possibility of adding the source of data, and the reliability the user puts in such source.

When data quality and validation are guaranteed then medical records might be more suitable data sources in clinical trials [6].

4.3 The variation of the clinical measures

Pagnacco et al. [32] described that measurement errors are mostly random in nature. In other words, assuming a random nature of measurement errors also assumes that the within-subjects variability, i.e. the variance of the results obtained from the same subject, is similar between the subjects examined. This assumption usually holds true in engineering, where the “subjects” are inanimate objects. However, when dealing with humans and clinical measures, this supposition assumption is rarely satisfied because the variability is caused not only by external random factors but also by the subjects conditions and reactions to random endogenous and exogenous stimuli. Thus, the researchers should have a particular attention to the reasons for such variation in clinical measurements obtained in clinical trials, mainly intra-patients [30].

Different healthcare professionals can get different results when acquiring data from a patient (e.g. vital signs, height, weight). The variability in patient’s measurements affects the ability to have reliable results. Most of this variability can be explained by the professional level of training and experience, and also by patient’s individual characteristics [28]. The adoption of EHR can support quality control procedures, providing definitions and giving oriented advice at the moment of care (e.g. how to collect the data, confounding factors, etc.) [51]. This would improve measuring standardization and bring benefits to clinical trials data quality. Nevertheless, a fundamental challenge persists: how to control data variability related to patients’ characteristics.

5 Clinical measures in reality

Nutrition assessment seeks to detect nutritional problems, collaborating for the promotion and recovery of health [26]. For example, the bioimpedance is a useful method to assess the body composition. However, it has positive and negative points: there are so many rules to the exam to be performed as the patient does not drink alcohol, caffeine and does not do exercise 24h before the exam; women cannot do the exam in menstrual period, and the exam must be done in fasting state (of water too), and the bladder must be empty before the exam [48].

Thus, some studies have been reported that is much variability among the bioimpedance results, mainly related to the use of equations without the actual knowledge and the hydration status of the patient [12,23]. Therefore, we can perceive the importance of the patient to follow the stipulated rules for clinical research exams. The EHR can help

the health professional choose the better way to assess the patients body composition, but patient's involvement in the study is fundamental as he/she needs to follow the project rules.

Another example is the assessment of the dietary intake; in this case, intrapersonal variability is very significant. The estimation of consumption based on just a few days of collection leads to critical failures in this context [44]. In other hands, the short time of food observation does not reflect the habitual intake [9]. For this, the number of the days of evaluation and the kind of instrument (24h-record; diary food, etc.) are significant tools to obtain accurate results. The Institute of Medicine (IOM) [18], which published the Dietary Reference Intakes (DRI), takes into account both the variability of the nutrient requirement in individuals and intrapersonal variability of intake. For its application, however, it is necessary to use values of intrapersonal variability, expressed by the intrapersonal standard deviation of ingestion of each nutrient, obtained in studies with the same population [26]. However, we cannot forget that the patients commitment to filling all food instrument and not lie about the dietary intake, is a fundamental point to obtain correct data about the dietary intake.

Similarly, the blood pressure measurement presents intrinsic variability [16]. Blood pressure measurement in some cases is still performed in a non-standardised way [24]; some factors as the health professional, environment; equipment, technical and the patient can interfere in trustworthiness to blood pressure assessment. The protocol recommendations include avoiding physical exercise 60 min before the exam, drinking alcohol, coffee, smoking 30 min before the assessment, not talking and keeping your legs uncrossed, these are just a few recommendations [34].

Besides, other types of patients intrapersonal variability, environments are important to the trustworthiness of clinical trials results. Rodriguez-Segade et al. [39] investigated the association between nephropathy and HbA1c variability in 2103 patients followed up for a mean 6.6 years. The authors concluded that in patients with type 2 diabetes, the risk of progression of nephropathy increases significantly with HbA1c variability, independently of updates mean HbA1c. To explain this point, the author hypothesizes that lifestyle influences HbA1c variability. Greater HbA1c variability having been reported to be associated with unfavourable lifestyle factors among patients with type 1 diabetes [50]. However, other authors demonstrate that this variability can be involved in a low socioeconomic class [50], and insulin resistance, which it has been implicated in the pathogenesis of diabetes complications [13]. These studies demonstrate that the patient's variability is associated with environments and also genetic factors which are essential be considered by the clinical trials investigators.

Another issue is clinical evaluator variability. To reduce evaluator's variability the same person should do all the clinical measures. But even then it still would be impossible to guarantee that health professionals are systematic in performing all tasks, i.e. positioning the cuff of the blood pressure device exactly in the same position every time, or taking the same amount of adipose panicle with the adipometer to assess body composition.

Thus, we understand that computer system can improve data quality, but unfortunately, the adequacy of the patient's clinical trial rules, and storing information about the protocol used in each case, are still essential for this quality of data.

6 Personalised medicine: pros and cons

Personalised medicine is defined as the use of genomic and other biotechnologies to derive information about an individual that could then be applied to obtain information on types of health interventions that would best suit that individual [4,41]. Over recent years, considerable technical advances have increasingly linked personalised medicine with preventive medicine. Although this process provides benefits in treating patients, particularly regarding the genetic profile, challenges, mainly regarding the lay public, still exist. Thus, certain points must be discussed to ensure the protection and fair treatment of individuals [20,42].

Personalised medicine is an example of what medicine desires to be in the future: specific, rigorous, and able to control disease and death [41]. For this, this field applies tools that enable risk assessment and prediction, such as health risk assessments, family history and, mainly, genetic information [11]. The advantages of personalised medicine are not simply applicable to patient treatment, but also aid in the prevention and prediction of disease by identifying genetic predispositions, predicting a potential patient [11,41].

Recent studies have been reported concerning the benefits of personalised medicine, which covers biomarker used to detect specific genetic traits and to guide different approaches towards the prevention and treatment of diseases, offering substantial healthcare savings [35]. Najafzadeh et al. [29] described several potential advantages of personalised medicine, such as possible applications of pharmacogenomics in tailoring treatments to improve effectiveness and minimise adverse effects; disease diagnosis; genomic testing in preventive medicine and the identification of new conditions.

In addition, cancer prevention and treatment appears as the greatest potential in the field of personalised medicine [15]. For example, regarding breast cancer, different immunologic markers have been applied to indicate the best treatment option and to assess metastasis and recurrence risk [1], whilst colon cancer therapy can be evaluated by genetic testing;

for example, homozygotes subjects for the UGT1A1, *28 allele show increased risk for neutropenia after treatment with irinotecan, with a reduction in the starting dose being advised [25].

However, although personalised medicine has been applied in multiple areas, mainly in oncology and cardiology, and many benefits for patient care have been noted, risks cannot be ignored. The field also offers many disadvantages, and several challenges are presents in this context, related to informed consent, confidentiality, genetic discrimination and direct-to-consumer genetic testing, among others [20].

Nevertheless, as research/treatment makes use of personalised medicine, becoming increasingly more common, the management of the ethical and legal issues becomes even more necessary. For example, in a hypothetical situation related to a patient that presented no response to a specific treatment, the health insurance requires genetic testing to determine drug safety and efficacy, in order to avoid unnecessary cost burdens. This point is an advantage regarding cost perspective, but it is not ethical, and the patient might prefer to incur the risks to avoid generating and releasing his/her genomic information [49]. Another ethical issue is related to genotype-driven research recruitment. This is a potentially powerful tool for studying the functional significance of human genetic variation. However, the genetic information generated for one study might be used as the basis for identifying and reconnecting participants for other studies [2]. These and other points justify the development of rules and clear guidelines to ensure the control and safety of the obtained information.

Other disadvantages and challenges in the use of personalised medicine can also be cited. Najafzadeh et al. [29] performed a semi-structured focus group with 28 physicians to discuss general themes about personalised medicine. From these focus groups, the authors categorised the disadvantages of personalised medicine expressed by experts into three perceived issues: validity uncertainty, equity issues, and implementation. The authors described that the physicians expressed concerns, mainly related to the uncertainty around the validity of genomic tests given the complexity of gene expression, which was mentioned as a major concern. Other points mentioned as potential disadvantages included financial incentives for private companies for excessive marketing of their services, possible mishandling of genomic information by private companies, and discrimination based on the genomic data (by insurance companies, the healthcare system, and employers).

In spite of the fact that personalised medicine considers the patient-centred approach to treatment and that it be most advantageous in the clinical practice, higher public engagement regarding this issue is required. Physicians also believe that lack of public knowledge about genomic tests presents possible unsafe impacts to patients after learning

about their predisposition to diseases, and the affordability of these tests for disadvantaged socioeconomic groups should also be taken into account [29]. In addition, citizens also appear concerned about personalised medicine perspectives; they believe that personalised tests might be used to ration care and that treatment should be applied only if the patient wants it. This issue raises clinical and policy challenges that may undermine the value of personalised medicine. Further efforts to deliberate with the public are warranted in order to inform effective, efficient, and equitable translations of personalised medicine [3].

Through these advantages, it is possible to understand that personalised medicine raises certain challenges, both for physician and healthcare systems, including the enormous number of available tests, the fast development of testing technologies, the decreasing unit cost per tested mutation and the potential of diagnostic and screening technologies to determine subsequent individual care pathways. Additionally, the support for required economic evidence to be produced to improve personalised medicine reimbursement and coverage decisions is also noteworthy [43].

From these described advantages and disadvantages, it is possible to understand the challenges of personalised medicine in order to aid in creating strategies to prepare the future healthcare system, reducing errors and improving the pros of personalised medicine.

7 Discussion

The quality of clinical data reducing errors of clinical results is a challenge. The electronic health record can help to improve it, but clinicians must have attention on innumerable factors, mainly related to patient involvement. We suggest that despite the use of EHR, health professionals should always discuss their results amply and attentively and promote the inclusion of the protocol used collecting data in an attempt to improve clinical data.

Information in healthcare institution has the potential to be very valuable due to the existing amount of data, and the economic value of the decisions they describe. To fulfil this potential today and in the future, it is critical that data scientists fully understand how these data were collected. Storing context information, protocols used and precision/accuracy information in clinical databases helps to ensure future understanding of such data.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Arpino, G., Generali, D., Sapino, A., Del Matro, L., Frassoldati, A., de Laurentis, M., Pronzato, P., Mustacchi, G., Cazzaniga, M., De Placido, S., et al.: Gene expression profiling in breast cancer: a clinical perspective. *Breast* **22**(2), 109–120 (2013)
- Beskow, L.M., Namey, E.E., Miller, P.R., Nelson, D.K., Cooper, A.: Irb chairs perspectives on genotype-driven research recruitment. *IRB* **34**(3), 1 (2012)
- Bombard, Y., Abelson, J., Simeonov, D., Gauvin, F.-P.: Citizens' perspectives on personalized medicine: a qualitative public deliberation study. *Eur. J. Hum. Genet.* **21**(11), 1197–1201 (2013)
- Burke, W., Psaty, B.M.: Personalized medicine in the era of genomics. *JAMA* **298**(14), 1682–1684 (2007)
- Cavallari, L.H., Sperber, N.R., Carpenter, J.S., et al.: Challenges and strategies for implementing genomic services in diverse settings: experiences from the implementing genomics in practice (ignite) network. *BMC Med. Genom.* **10**(1), 35 (2017)
- Cowie, M.R., Blomster, J.I., Curtis, L.H., Duclaux, S., Ford, I., Fritz, F., Goldman, S., Janmohamed, S., Kreuzer, J., Leenay, M., et al.: Electronic health records to facilitate clinical research. *Clin. Res. Cardiol.* **106**(1), 1–9 (2017)
- Cruz-Correia, R.J., Wyatt, J., Dinis-Ribeiro, M., Costa-Pereira, A.M.: Determinants of frequency and longevity of hospital encounters' data. *BMC Med. Inform. Decis. Mak.* **10**(1), 15 (2010)
- Cruz-Correia, R., Boldt, I., Lapão, L., Santos-Pereira, C., Rodrigues, P.P., Ferreira, A.M., Freitas, A.: Analysis of the quality of hospital information systems audit trails. *BMC Med. Inform. Decis. Mak.* **13**(1), 84 (2013)
- Cuppari, L.: Aplicacoes das dris na avaliacao da ingestao de nutrientes para individuos. In: Usos e aplicaes das dietary reference Inatakes DRI, Brasil I (2001)
- Eurostat: Mortality and life expectancy statistics, 2017
- Ginsburg, G.S., Willard, H.F.: Genomic and personalized medicine: foundations and applications. *Transl. Res.* **154**(6), 277–287 (2009)
- Graves, J.E., Pollock, M.L., Colvin, A.B., Van Loan, M., Lohman, T.G.: Comparison of different bioelectrical impedance analyzers in the prediction of body composition. *Am. J. Hum. Biol.* **1**(5), 603–611 (1989)
- Groop, P.-H., Forsblom, C.: Mechanisms of disease: pathway-selective insulin resistance and microvascular complications of diabetes. *Nat. Rev. Endocrinol.* **1**(2), 100 (2005)
- Harvey, A., Brand, A., Holgate, S.T., Kristiansen, L.V., Lehrach, H., Palotie, A., Prainsack, B.: The future of technologies for personalised medicine. *New Biotechnol.* **29**(6), 625–633 (2012)
- Hayes, D.F., Markus, H.S., Leslie, R.D., Topol, E.J.: Personalized medicine: risk prediction, targeted therapies and mobile health technology. *BMC Med.* **12**(1), 37 (2014)
- Imbelloni, L.E., Beato, L., Tolentino, A.P., de Souza, D.D., Cordeiro, J.A.: Monitores automáticos de pressão arterial: avaliação de três modelos em voluntárias. *Rev. Bras. Anestesiol.* **54**(1), 43–52 (2004)
- Imran, M., Hlavacs, H., Haq, I.U., Jan, B., Khan, F.A., Ahmad, A.: Provenance based data integrity checking and verification in cloud environments. *PLoS ONE* **12**(5), e0177576 (2017)
- Institute of Medicine: Dietary Reference Intakes: Applications in Dietary Assessment. National Academy Press, Washington (2000)
- Jansen, A.C.M., van Aalst-Cohen, E.S., Hutten, B.A., Büller, H.R., Kastelein, J.J.P., Prins, M.H.: Guidelines were developed for data collection from medical records for use in retrospective analyses. *J. Clin. Epidemiol.* **58**(3), 269–274 (2005)
- Joly, Y., Saulnier, K.M., Osien, G., Knoppers, B.M.: The ethical framing of personalized medicine. *Curr. Opin. Allergy Clin. Immunol.* **14**(5), 404–408 (2014)
- Jung, M.-H., Kim, G.-H., Kim, J.-H., Moon, K.-W., Yoo, K.-D., Rho, T.-H., Kim, C.-M.: Reliability of home blood pressure monitoring: in the context of validation and accuracy. *Blood Press. Monit.* **20**(4), 215–220 (2015)
- Kuhle, S., Maguire, B., Ata, N., MacInnis, N., Dodds, L.: Birth weight for gestational age, anthropometric measures, and cardiovascular disease markers in children. *J. Pediatr.* **182**(21), 99–106 (2016)
- Kushner, R.F.: Bioelectrical impedance analysis: a review of principles and applications. *J. Am. Coll. Nutr.* **11**(2), 199–209 (1992)
- Lehane, A., O'Brien, E.T., O'Malley, K.: Reporting of blood pressure data in medical journals. *Br. Med. J.* **281**(6255), 1603 (1980)
- Liu, C.-Y., Chen, P.-M., Chiou, T.-J., Liu, J.-H., Lin, J.-K., Lin, T.-C., Chen, W.-S., Jiang, J.-K., Wang, H.-S., Wang, W.-S.: Ugt1a1*28 polymorphism predicts irinotecan-induced severe toxicities without affecting treatment outcome and survival in patients with metastatic colorectal carcinoma. *Cancer* **112**(9), 1932–1940 (2008)
- Marchioni, D.M.L., Junior, E.V., Galvão Cesar, C.L., Fisberg, R.M.: Avaliação da adequação da ingestão de nutrientes na prática clínica. *Rev. Nutr.* **24**(6), 825–832 (2011)
- McColl, E., et al. Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. Core Research (2001). https://www.researchgate.net/profile/Lois_Thomas/publication/11550481_Design_and_use_of_questionnaires_a_review_of_best_practice_applicable_to_surveys_of_Health_Service_staff_and_patients/links/00b7d52df7cdf18db6000000/Design-and-use-of-questionnaires-a-review-of-best-practice-applicable-to-surveys-of-Health-Servicestaff-and-patients.pdf
- McGregor, M., Cambron, J.A., Jedlicka, J., Gudavalli, M.R.: Clinical trial variability: quality control in a randomized clinical trial. *Contemp. Clin. Trials* **30**(1), 20–23 (2009)
- Najafzadeh, M., Davis, J.C., Joshi, P., Marra, C.: Barriers for integrating personalized medicine into clinical practice: a qualitative analysis. *Am. J. Med. Genet. Part A* **161**(4), 758–763 (2013)
- Normando, D.: Dental press journal of orthodontics: one year later, and more growth. *Dent. Press J. Orthod.* **22**, 9–10 (2017)
- Novaes, J.F., Lamounier, J.A., Colosimo, E.A., Franceschini, S.C.A.: Breastfeeding and obesity in Brazilian children. *Eur. J. Public Health* **22**(3), 383–389 (2012)
- Pagnacco, G., Carrick, F.R., Wright, C.H.G., Oggero, E.: Between-subjects differences of within-subject variability in repeated balance measures: consequences on the minimum detectable change. *Gait Posture* **41**(1), 136–140 (2015)
- Pasquier, T., Lau, M.K., Trisovic, A., Boose, E.R., Couturier, B., Crosas, M., Ellison, A.M., Gibson, V., Jones, C.R., Seltzer, M.: If these data could talk. *Sci. Data* **4**, 170114 (2017)
- Pickering, T.G., Hall, J.E., Appel, L.J., Falkner, B.E., Graves, J., Hill, M.N., Jones, D.W., Kurtz, T., Sheps, S.G., Roccella, E.J.: Recommendations for blood pressure measurement in humans and experimental animals. *Circulation* **111**(5), 697–716 (2005)
- Pinho, J.R.R., Sitnik, R., Mangueira, C.L.P.: Personalized medicine and the clinical laboratory. *Einstein (São Paulo)* **12**(3), 366–373 (2014)
- Rossi, S., et al.: p-Medicine: from data sharing and integration via VPH models to personalized medicine. *Ecancermedscience* (2011). <https://doi.org/10.3332/ecancer.2011.218>

37. Project DISCIPULUS: Digitally integrated scientific data for patients and populations in user-specific simulations—project id: 288143, 2013
38. Project myhealthavatar: A demonstration of 4d digital avatar infrastructure for access of complete patient information—project id: 600929, 2013
39. Rodriguez-Segade, S., Rodriguez, J., Garca Lpez, J.M., Casanueva, F.F., Camia, F.: Intrapersonal hba1c variability and the risk of progression of nephropathy in patients with type 2 diabetes. *Diabetic Med.* **29**(12), 1562–1566 (2012)
40. Saenger, P., Czernichow, P., Hughes, I., Reiter, E.O.: Small for gestational age: short stature and beyond. *Endocr. Rev.* **28**(2), 219–251 (2007)
41. Savard, J.: Personalised medicine: a critique on the future of health care. *J. Bioethical Inquiry* **10**(2), 197–203 (2013)
42. Scheen, A.J., Giet, D.: Personalized medicine: all benefits for the patient but new challenge in the physician–patient relationship. *Rev. Med. Liege* **70**(5–6), 247–250 (2015)
43. Schleidgen, S., Klingler, C., Bertram, T., Rogowski, W.H., Marckmann, G.: What is personalized medicine: sharpening a vague term based on a systematic literature review. *BMC Med. Ethics* **14**(1), 55 (2013)
44. Sempos C.T., Looker A.C., Johnson C.L., Woteki C.E. The importance of within-person variability in estimating prevalence. In: Macdonald, I. (eds.) *Monitoring dietary intakes*. ILSI Monographs. Springer, London (1991). https://doi.org/10.1007/978-1-4471-1828-2_9
45. Shortliffe, E.H., Cimino, J.J.: *Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics)*, 4th edn. Springer, Berlin (2015)
46. Strickland-Marmol, L.B., Muro-Cacho, C.A., Barnett, S.D., Banas, M.R., Foulis, P.R.: College of American pathologists cancer protocols: optimizing format for accuracy and efficiency. *Arch. Pathol. Lab. Med.* **140**(6), 578–587 (2016)
47. Kalra, D., Beale, T., Heard, S.: The openEHR foundation. *Stud health Technol Inform* **115**, 153–173 (2005)
48. Vitolo, M.R. Avaliação nutricional no adulto. In: Vitolo, M.R. *Nutrição: da gestação ao envelhecimento*, 2nd edn, pp. 377–397. Rubio, Rio de Janeiro (2008)
49. Vogenberg, F.R., Barash, C.I., Pursel, M.: Personalized medicine: part 2: ethical, legal, and regulatory issues. *Pharm. Ther.* **35**(11), 624 (2010)
50. Wadén, J., Forsblom, C., Thorn, L.M., Gordin, D., Saraheimo, M.: A1c variability predicts incident cardiovascular events, microalbuminuria, and overt diabetic nephropathy in patients with type 1 diabetes. *Diabetes* **58**(11), 2649–2655 (2009)
51. Weiskopf, N.G., Weng, C.: Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J. Am. Med. Inform. Assoc.* **20**(1), 144–151 (2013)
52. Wyatt, J.C., Wright, P.: Design should help use of patients' data. *Lancet* **352**, 1375–1378 (1998)
53. Xiao, L., Cousins, G., Courtney, B., Hederman, L., Fahey, T., Dimitrov, B.: Developing an electronic health record (EHR) for methadone treatment recording and decision support. *BMC Med. Inform. Decis. Mak.* **11**(5), 2–10 (2011)
54. Zarrati, M., Shidfar, F., Moradof, M., Nejad, F.N., Keyvani, H., Hemami, M.R., Razmpoosh, E.: Relationship between breast feeding and obesity in children with low birth weight. *Iran. Red Crescent Med. J.* **15**(8), 676–682 (2013)