Giuseppe Banfi

Big Data-Big Opportunity for Patient Associations and Clinical Trial Center

Big data, differently from what many believe, do not regard just the merge of many registries and clinical Databases, nor the establishment of new unified single repositories, into which clinicians and their institutions can transmit large sets of structured data, collected by means of standardized forms. Big data, by definition, regard: very *large amount* of data (which could only be partially the case of many forms collected together); data that change or are acquired *quickly* (not only few times a day or weeks); and most importantly a great *variety* of data. This means that structured data are only one small portion of the data that people call big; a common place in the IT discourse runs that unstructured (i.e. untyped data) data are the "*other* 80 percent of *all* data": in other words, it is unstructured whatever is *not* in a databases. Unstructured data usually encompass sensor data; text-based practice guidelines and health product information; scientific publications and case reports; as well as (health-related) content produced by people within their favorite social media (also unaware of doing it); clinical imaging resources; and, last but not the least, the free-text areas and comments contained in any electronic medical record.
The emphasis (and economic investment) on the big data can, almost unintentionally, allow for a renovated appreciation of the "small data", especially the totally unstructured ones, that are produced, annotated and consumed locally at the doctor office.
The latest achievements in data mining and pattern recognition do have a potential to improve the doctors' ability to leverage the patients' accounts of their history and symptoms, and it could be included in secondary outcomes of clinical trials.

Personalized medicine aims to trace medical evidences an recommendations back to the "particular person" instead of considering average classes of "similar" patients (cf. the diagnosis-related groups introduced in the medical discourse 35 years ago). This medicine is enabled by big data analytics and current data mining and pattern recognition techniques when these are coupled with a comprehensive datafication of the patient characteristics.

Personalized medicine absolutely needs big data. The "in one clinical trials" need big data to achieve success.

EU in proposing "big data" as a horizontal theme for all biomedical topics. Big data should be fundamental for improving clinical trials, especially to avoid unnecessary experimental phases, to cut long times elapsed from the start of recruitment of patients and first elaboration of data, and to emphasize "in silico trails".

Moreover, big data should be crucial for evaluating continuous biomedical data and findings, e.g. by point of care or wearable sensors, and for mimicking "real-life"

conditions during the experimental phase for drugs, medical devices, probiotics and procedures.

Big data could be obtained by laboratory medicine and additional diagnostic repositories (imaging): the standardization of procedures to achieve data is mandatory. Living lab could be a common place for hospital experimental trials.

We could present some experiences of San Raffaele and Galeazzi, concerning/including robotics, PROMS, registries, role of nurses