Online tool for Massive Observational Studies (OMOS) in Drug Repositioning and Cancer Prevention: A Technical Overview

Phung-Anh (Alex) Nguyen, Ph.D.[1], Usman Iqbal, PharmD, MBA[1], Chih-Wei Huang, MS[1], Shen-Hsien Lin, MS[1], Yu-Chuan (Jack) Li, M.D., Ph.D. [1,2,*]

[1]College of Medical Science & Technology, Taipei Medical University, Taiwan;

[2]Department of Dermatology, Wan-Fang Hospital, Taiwan; * Corresponding author

Abstract

Health observational data also known as big-data is leveraging every day. To utilize this in effective way for generating health outcomes of interest specifically. We developed a model which is efficient and cost effective to do mass production studies. Methodology designed with experts help based upon existing statistical methods for pharmacoepidemiology studies. Model help to generate automatically case-controls by using large databases, to do mass production studies, saving time, cost effective and don't require professionals.

Introduction

Pre-marketing drug evaluation processes are still largely based on formal, extremely expensive randomized clinical trials and manual information collection processes that cover a relatively small sample of patients.[1] Moreover, the rapid change in health information technology system had dramatically increased health data accumulated.[2] It has become an important material that provides an extraordinary opportunity to observe the emergence of new knowledge and its influence, particularly in drug repositioning and cancer prevention.[3] Upon this beneficial, we aimed to develop an online informatics tool in order to evaluate the risk of drugs for cancer by utilizing medical big data. It could produce the massive studies, reduce cost, time verses traditional trials, and help to improve drug safety, quality in health care.

Methods

Data Source: We use the Taiwan's National Health Insurance Database that has provided a huge data which covered all health information including characteristics and all drug information i.e. prescriptions, etc. of 23 million Taiwanese population.

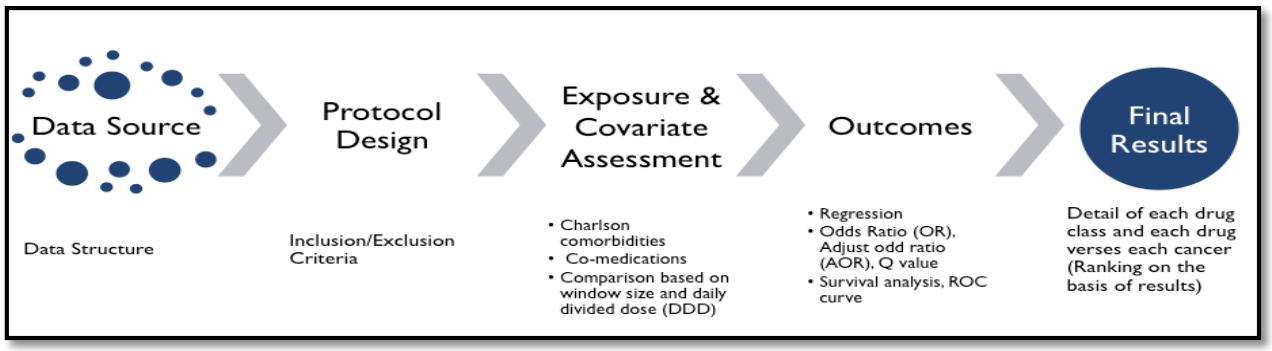

Figure 1. The processing of OMOS system

System overview: The processing of OMOS system is consists of five steps, which are shown in Figure 1.

Front-end development: Web-based interface was developed by using PHP package and Javascript. In addition, we included the guidelines of evidence based medicine (EBM) level 3 for observational study such as cohort, case-control, and/or case serial self-control in order to support users interact with system.[4]

Back-end development: A package of Apache, MySQL & PHP was used to build the serve-side of the system. We integrated the Elasticsearch API[5] to our system in order to search and analyze data immediately. The example of data transform to person-level from Taiwan NHI database is shown in Box 1. After then, we also integrated the analytics package (ie. R package) to perform the statistical analysis to a given study.

Results

We successfully developed and built the informatics tool, which used large observational health data to study the effects of medical interventions and predict health outcomes, especially for long-term use drug verse cancer risk. Our system has been design to satisfy few requirements such as mass production studies, save time, cost effectiveness, and do not require professionals. The UI of our system is shown in Figure 2.

Box 1. An example of data transform from Taiwan's NHI claim database

```
"cd": { //Records of diseases
"properties": {
        "birthday_year": {"type": "integer"},
        "gender": {"type": "string"},
        "pid": {"type": "string"},
        "sn": {"type": "long"},
        "subtotal": {"type": "nested",
                "properties": {
                "count":{"type": "integer"},
                "first_days":{"type": "integer"},
                "icd9": {"type": "string"},
                "last_days": {"type": "integer"}
                }
        }
}
}
```

```
"oo_drug": { //Records of drugs use
"properties": {
        "birthday_year": {"type": "string"},
        "gender": {"type": "string"},
        "oo_drug": {"type": "nested",
                "properties": {
                "atc_code": {"type": "string"},
                "days": {"type": "integer"},
                "drug_day": {"type": "integer"},
                "drug_no": {"type": "string"},
                "drug_use": {"type": "integer"}
                }
        },
        "pid": {"type": "string"},
        "sn": {"type": "long"}
}
}
```

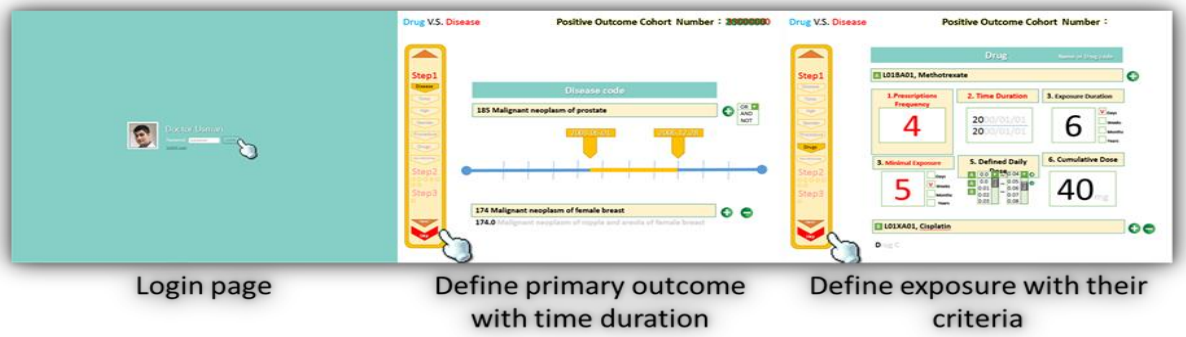

Figure 2. An example of UIs in the system

Discussions and Conclusions

Developing a model which is capable for long term drug use and cancer risk on a societal scale is a big challenge that is approachable, achievable, and has implications towards those developing and/or using medications. Such research model would also provide an excellent test bed for solving the technological, informatics, and organizational issues towards other broad domains of drug evaluation mass production studies by utilizing large-scale databases.

References

1. Hiatt WR. Observational Studies of Drug Safety — Aprotinin and the Absence of Transparency. *New England Journal of Medicine.* 2006;355(21):2171-2173.
2. Blumenthal D, Tavenner M. The "Meaningful Use" Regulation for Electronic Health Records. *New England Journal of Medicine.* 2010;363(6):501-504.
3. Benson K, Hartz AJ. A Comparison of Observational Studies and Randomized, Controlled Trials. *New England Journal of Medicine.* 2000;342(25):1878-1886.
4. Doi SAR. *Understanding evidence in health care: Using clinical epidemiology.* South Yarra, VIC, Australia: Palgrave Macmillan; 2012.
5. BV E. Elasticsearch. https://www.elastic.co/. Accessed Sep 07, 2015.