

Assessing the Collective Population Representativeness of Related Type 2 Diabetes Trials by Combining Public Data from ClinicalTrials.gov and NHANES

Zhe He^a, Shuang Wang^b, Elhaam Borhanian^a, and Chunhua Weng^a

^aDepartment of Biomedical Informatics, Columbia University, New York, NY, USA

^bDepartment of Biostatistics, Columbia University, New York, NY, USA

Abstract

Randomized controlled trials generate high-quality medical evidence. However, the use of unjustified inclusion/exclusion criteria may compromise the external validity of a study. We have introduced a method to assess the population representativeness of related clinical trials using electronic health record (EHR) data. As EHR data may not perfectly represent the real-world patient population, in this work, we further validated the method and its results using the National Health and Nutrition Examination Survey (NHANES) data. We visualized and quantified the differences in the distributions of age, HbA1c, and BMI among the target population of Type 2 diabetes trials, diabetics in NHANES databases, and a convenience sample of patients enrolled in selected Type 2 diabetes trials. The results are consistent with the previous study.

Keywords:

Clinical Trial; Patient Selection; Selection Bias.

Introduction

Randomized controlled trials are the gold standard for producing high-quality medical evidence. However, they may suffer from delayed enrollment and lack of population representativeness, resulting in compromised generalizability to the real-world patients to whom the results or findings of a trial are usually applied. To assess the external validity of a trial, researchers often compared the study population of its enrolled patients and the patient population with certain medical conditions, e.g., breast cancer [1] or major depression [2]. Our previous work showed that many clinical trials, especially those on the same medical condition, use similar or identical eligibility criteria [3]. Therefore, we hypothesize that the generalizability issue might be at not only the individual trial level but also the community level of the entire clinical trial enterprise.

As a step to advance the field for generalizability assessment, we previously proposed a method to compare patient populations in electronic health records (EHRs) with aggregated clinical trial target populations that characterize the patients who can be recruited in a set of trials according to their inclusion and exclusion criteria [4]. This method also introduced *Generalizability Index for Study Traits (GIST)* for quantifying the population representativeness of clinical trials to the general patient population. This approach is advantageous over existing methods [1, 2] in that generalizability assessment can be performed proactively during design. Using this method, we found that Type 2 diabetes studies are more generalizable with respect to age than they are with respect to HbA1c. However, one of the limitations of the study is the potential bias in EHR data towards certain population subgroups [5].

The available public datasets offer great opportunity to further validate the method and the results generated from EHR data.

The National Health and Nutrition Examination Survey (NHANES) is a continuous cross-sectional health survey conducted by the National Center for Health Statistics of CDC [6]. It evaluates a stratified multistage probability sample of the non-institutionalized population of the United States. The survey samples are first interviewed at home, followed by a physical and laboratory test in a mobile examination center. Its rigorous quality control standards ensure national representativeness and high-quality data collection. NHANES data have facilitated various translational bioinformatics research. Chen *et al.*, for example, have built a predictive aging model of adolescent development [7]. Bays *et al.* have analyzed connections between body mass index and metabolic diseases [8]. These promising results have propelled our use of NHANES data to support our population-based study. In this paper, we assess the collective population representativeness of multiple related Type 2 diabetes mellitus (T2DM) trials through pairwise comparison between the patient population derived from the NHANES data, the target population derived from clinical trial summaries, and a convenience sample of study population derived from the results of selected trials in ClinicalTrials.gov. We hypothesize that NHANES data can be used to assess the population representativeness of trials and to validate the results of our previous study using EHR data [4].

Methods

Figure 1 shows the workflow of this study. We first identified frequently used quantitative eligibility features (i.e., with a permissible value range) in T2DM trials between 2003 and 2012. In the NHANES data of continuous survey years 2003 to 2012, we identified sample subjects with T2DM and extracted interview and laboratory test results relevant to this work. We also extracted results of selected T2DM clinical trials between 2003 and 2012 from ClinicalTrials.gov and used these data to profile a convenience sample of the study population. We visualized and analyzed the differences among the T2DM patient population, the target populations of T2DM trials, and a convenience sample of study populations of enrolled patients in T2DM trials, all of which were profiled by one eligibility feature at a time. Then we calculated the GIST scores [4] for each eligibility feature.

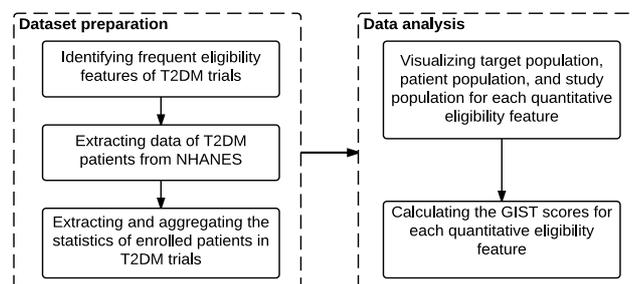


Figure 1– The workflow of the study

Dataset Preparation

Processing clinical trial summaries

We have built and published a database COMPACT [9], which stores parsed eligibility criteria and trial descriptors of 162,586 clinical study summaries in ClinicalTrials.gov downloaded on March 18, 2014. All the trials were indexed by medical conditions using the API provided by ClinicalTrials.gov. From COMPACT, we retrieved 2,731 interventional studies on T2DM with a start date between 01/2003 and 12/2012. Age, HbA1c, and body mass index (BMI) were identified as three most frequently used quantitative eligibility features that appear in 2,702 (99.0%), 1,463 (53.6%), and 1,274 (46.6%) of the trials, respectively. We formed one trial set for each of the three features and included them for analyzing the distribution of trials over age, HbA1c, and BMI, respectively.

Processing NHANES data

We extracted relevant interview data and laboratory measurements for each NHANES sample during 2003 – 2012. The data elements included *SEQN* (the unique identifier of a sample), *gender*, *age*, *race/ethnicity*, *BMI*, *Glycohemoglobin (HbA1c) values*, *two-year sample weights for the interview and mobile examination portion of the study*, *two year cycle*, and interview questions “*Were you told to have diabetes by a doctor or a health professional*”, “*Age when first you were told to have diabetes*”, and “*Are you taking insulin?*” The laboratory methodology to measure Glycohemoglobin value of NHANES sample subjects can be found at [10].

In NHANES, we identified 3,304 diabetic sample subjects who were told to have diabetes by a doctor or a health professional and had an HbA1c measurement [11]. Since NHANES participants were not asked to report the type of diabetes (type 1 or type 2) they were diagnosed with, we employed a method used by Dodd *et al.* [11] to further exclude 222 samples with Type 1 diabetes who were (1) first diagnosed with diabetes before age 30; and (2) taking insulin. The rationale is that as one grows older, his/her lifestyle (e.g., dietary habits) will play a more important factor in developing T2DM; thus, if a person is diagnosed with diabetes at a young age and takes only insulin, it is likely that s/he is with Type 1 diabetes.

Out of the 3,082 T2DM samples, 2,695 had no missing values for age, HbA1c, and BMI. Because the number of samples with missing data (3,082 - 2,695) exceeded 10% of the total number of T2DM samples, we used two categorical variables gender and race/ethnicity that every sample of NHANES has to assess the representativeness of these 2,695 samples for all 3,082 T2DM samples. With respect to gender and race/ethnicity, we used the chi-square test to test the statistically significant differences between any pair of the following three sets of samples: (1) all 3,082 samples, (2) 2,695 samples with age, HbA1c and BMI values, and (3) 387 samples with missing values. No statistically significant pair-wise difference was found between any pair of samples (all *p-values* > 0.05). Further, we used the two-sample *t*-test to test the pair-wise differences for HbA1c, age, BMI. No statistically significant difference was found (all *p-values* > 0.05). Therefore, we concluded that the 2,695 samples that had complete age, HbA1c, and age values is a representative sample of all the T2DM samples in NHANES and we used this sample as the patient cohort of this study.

To account for complex survey design (e.g., oversampling), non-response, and post-stratification, NHANES assigned each sample subject a sample weight, which represents the number of people in the U.S. national population that a specific sample can represent. When a sample is weighted, it is representative of the U.S. civilian non-institutionalized Census population

[12]. NHANES publishes survey data once every two years. In this study, we combined data of five two-year cycles from 2003 to 2012. Following the analytical guideline of NHANES [13], we used WTMEC2YR as the sample weight and calculated the ten-year sample weight WTMEC10YR (1/5 * WTMEC2YR). After taking the ten-year sample weight into account, these 2,695 samples can thus represent 15,575,484 T2DM patients in the U.S. national population. More importantly, the distribution of patients with sample weight can represent the real distribution of the U.S. national population. In this paper, all the subsequent analyses were performed after taking sample weights into account. Table 1 shows the baseline characteristics of the T2DM patient cohort of the study.

Table 1– Baseline characteristics of patient cohort.

Characteristics	Number
Sample, n	2,695
Population, n	15,575,484
Age (mean ± SD)	60.6 ± 13.3
Gender, male%	48.2
Race/Ethnicity	
Mexican American, %	8.5
Other Hispanic, %	5.5
Non-Hispanic White, %	61.8
Non-Hispanic Black, %	16.0
Other races, %	8.2
HbA1c, % (mean ± SD)	7.2 ± 1.7
BMI, kg/m ² (mean ± SD)	32.9 ± 7.6

Processing clinical trial results

To profile the study populations, we parsed the baseline characteristics of enrolled patients for the T2DM studies that published results in ClinicalTrials.gov. Out of 2,731 T2DM trials between 2003 and 2012, only 531 reported their baseline characteristics of enrolled participants in ClinicalTrials.gov. The numbers of trials that reported mean and standard deviation (SD) of age, HbA1c, and BMI of their enrolled participants are 389, 137, and 108, respectively. Since only a small portion of trials reported results in ClinicalTrials.gov, the study population included in this analysis represents merely a convenience sample of the study population. For each study, we extracted participant counts, mean and SD for age, HbA1c, and BMI. We aggregated the mean and SD for each feature using the following formulas (adapted from [14]), where *T* is the number of studies,

$$\text{Weighted_mean} = \frac{\sum_{i=1}^T (\text{mean}_i * \text{number_participants}_i)}{\sum_{i=1}^T \text{number_participants}_i} \quad (1)$$

$$\text{Weighted_SD} = \sqrt{\frac{\sum_{i=1}^T (\text{SD}_i^2 * (\text{number_participants}_i - 1))}{\sum_{i=1}^T (\text{number_participants}_i - 1)}} \quad (2)$$

Visualizing populations

Effective visualization of target populations, patient population, and study population over a quantitative trait allows interested viewers to easily discern the differences among them. To profile the target populations of T2DM studies represented by a quantitative eligibility feature, we plotted the distribution of trials over a quantitative feature (i.e., age, HbA1c, and BMI). This distribution shows the number of trials that recruit patients with a certain value, which can reveal systematically excluded or overly included value ranges. The patient population of T2DM was presented by the distributions of several patients' features. The study population was similarly plotted using Gaussian distributions with weighted means and SDs of several features.

For each feature, we plotted the distributions of patient population, target population, and study population in the same

figure. We employed 2nd degree polynomial local weighted regression fitting with a span of 20% to smooth the curves [15].

Quantitative metric of population representativeness

To quantify the population representativeness of studies for a single quantitative feature, such as age, HbA1c, and BMI, we calculated the GIST scores [4]. GIST is the sum across all consecutive non-overlapping value intervals of the percentage of studies that recruit patients in that interval, multiplied by the percentage of patients observed in that interval:

$$GIST = \frac{\sum_{i=1}^N \sum_{j=1}^T I([i_{low}, i_{high}] \subset w_j)}{T} * \frac{\sum_{k=1}^P I(i_{low} \leq y_k < i_{high})}{P} \quad (3)$$

where N is the number of distinct value intervals of the quantitative feature, T is the number of trials, P is the number of patients, w_j is the inclusion value interval of the quantitative feature for the j^{th} study, such that indicator I can be defined as j^{th} study interval subsumes the i^{th} interval low and high boundary values, and y_k is the observed value of the quantitative feature for the k^{th} patient such that an indicator I can be defined when k^{th} patient has a value of the quantitative feature falls within the i^{th} interval.

The GIST score ranges between 0 and 1 and characterizes the proportion of patients that would be potentially eligible across trials, with 1 being perfectly generalizable and 0 being completely not generalizable. Note that the formula for calculating the GIST score can also be applied to categorical variables, whereby the values are integers.

Results

Visualization of populations

Figure 2 shows the juxtaposition of the distributions of T2DM patients (blue solid curve), T2DM trials (green dot-and-dashed curve), and study samples enrolled in T2DM trials (red dashed curve) over age values. The x-axis represents the age values. The left y-axis represents the percentage of patients of a certain age. The right y-axis represents the percentage of trials that recruit patients of a certain age. Note that both target population (green dot-and-dashed curve) and study population (red dashed curve) use the same value ticks on the right y-axis. From the visualization, we can observe that T2DM trials recruit patients with broad age range (between 18 – 80), whereas most real-world patients are older. Comparing the target population (green dot-and-dashed curve) and study population (red dashed curve), we can see that even though T2DM studies tend to recruit patients of broad age range, most enrolled patients were between 40 to 70 years old.

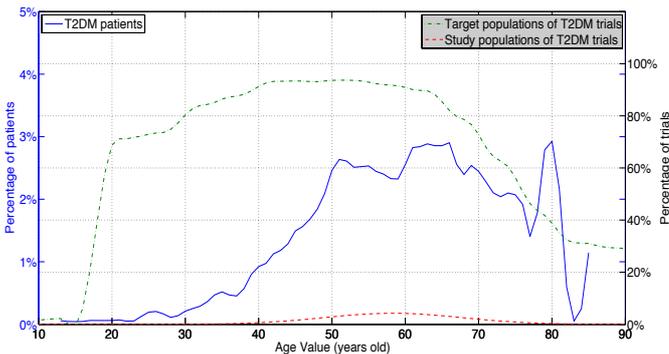


Figure 2– Target population, patient population, and study population of T2DM trials on age values

Figure 3 visualizes the comparison of these three populations over HbA1c values. Between years 2003 and 2012, targeted HbA1c values in T2DM trials are consistently higher than that of the patient population, indicating that T2DM trials tend to recruit sicker diabetic patients. This observation is consistent with the results of our previous study [4]. There is also a noticeable shift between the study population (red dashed curve) and the patient population (blue solid curve), indicating that on average patients enrolled in T2DM trials had a higher HbA1c value than the general T2DM patient population. The curves for study patients and target patients aligned well with similar curve shapes.

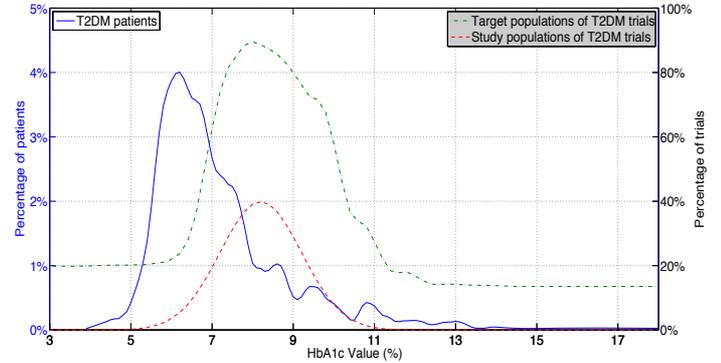


Figure 3– Target population, patient population, and study population of T2DM trials on HbA1c values

Figure 4 is the visualization for BMI. We can see that three curves aligned better than those for age and HbA1c, indicating that T2DM studies may have a better population representativeness with respect to BMI. The concentration of BMI values is accordance with real-world T2DM patients.

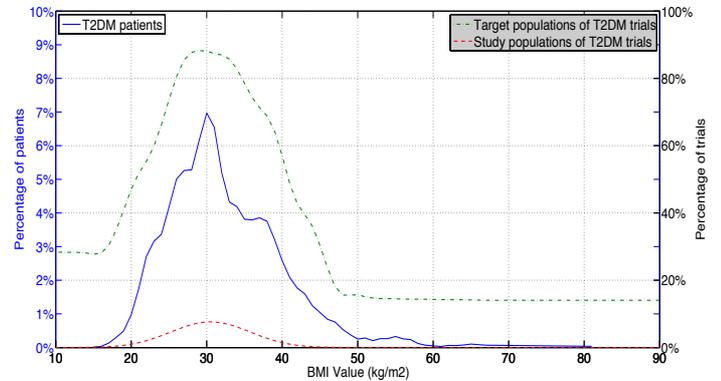


Figure 4– Target population, patient population, and study population of T2DM trials on BMI values

Comparison: study population vs. patient population

Table 2 presents the comparison between the study population (i.e., patients who were enrolled in T2DM trials) and the general patient population. One-sample t -test was performed to test the statistical significance of the differences between these two populations for each feature, whose mean value of the general patients was used as hypothetical mean. As shown in the table, statistically significant differences between the study population and the general patient population were observed for all three features. On average, the patients enrolled in T2DM trials are younger, with lower BMI and higher HbA1c than the general patient population ($p < 0.0001$).

Table 2– Comparison between the convenience sample of the study population and the patient population.

Feature (mean \pm SD)	Study population	Patient population	Difference, mean (95% CI)	P value
N	198,050	15,575,484	--	--
Age	58.3 \pm 9.4	60.6 \pm 13.3	- 2.3 (- 2.34 to - 2.26)	< 0.0001
N	62,931	15,575,484	--	--
HbA1c	8.2 \pm 1.0	7.2 \pm 1.7	1.0 (0.99 to 1.01)	< 0.0001
N	70,678	15,575,484	--	--
BMI	30.5 \pm 5.2	32.9 \pm 7.6	- 2.4 (- 2.44 to - 2.36)	< 0.0001

GIST scores

The GIST scores were computed for age, HbA1c, and BMI for each phase using fixed-width value intervals of width 1.0, 0.5, and 1.0, respectively (Table 3). Note that large value intervals may exclude some trials; thus we used these small interval widths to ensure that all the trials were included. To calculate the GIST score correctly, one could also use all the threshold values of a feature in all the trials to divide its value spectrum into bins of varying widths. The overall GIST scores show that T2DM trials have the best population representativeness with regard to age, followed by BMI and then HbA1c. Comparing the results across trials in different phases, we observed that the GIST of age increases from Phase I to Phase III, whereas the GIST of HbA1c decreases from Phase I to Phase III, which confirms our previous study using EHR data [4].

We have provided the interim data and results of this work as supplementary material, which can be accessed at <http://is.gd/tbwzj9>.

Table 3– GIST scores of age, HbA1c, and BMI for each phase.

GIST Phase	Variable		
	Age	HbA1c	BMI
I	0.60 (N=368)	0.55 (N=141)	0.64 (N=204)
II	0.77 (N=517)	0.45 (N=244)	0.71 (N=194)
III	0.87 (N=766)	0.38 (N=438)	0.77 (N=356)
IV	0.80 (N=484)	0.42 (N=306)	0.69 (N=194)
All	0.77 (N=2702)	0.44 (N=1463)	0.69 (N=1274)

Discussion

In this study, we used both visualization and quantitative metric GIST to assess the population representativeness of T2DM trials to the general T2DM patient population. One interesting observation is that even though the curves of BMI (Figure 2) aligned better than those of age (Figure 4) in the visualization, the GIST score of BMI (0.69) is lower than age (0.77), showing that visualization may not be sufficient for comparing population representativeness between variables. Conversely, the GIST score may not be as intuitive as visualization for discerning systematic included or excluded value ranges.

Clinical trials may serve varying purposes. Not all clinical trials need to generalize to a broad patient population. Many trials require a “clean” and specific cohort to ensure internal validity, which is usually achieved by minimizing confounding variables with tightened eligibility criteria. Therefore, the purpose of this study is not to demonstrate the systematic bias in T2DM trials, but to uncover collective design patterns via

aggregate analysis of multiple trials. We identified and visualized the significant differences among the aggregated study populations, target populations in trials that investigate the same medical condition, and patient populations that are supposed to benefit from these trials. Our method intends to quantify the differences among these three populations so that they can differentiate them for research in any disease domain and make informed decisions for future trial designs.

NHANES data vs. EHR data

Using NHANES data, we generated results that are consistent with our previous study using EHR data [4]. We further validated the effectiveness of the method for assessing the population representativeness of clinical trials. With the large amount of patient data existing in both EHRs and public databases, it is important to understand their respective strengths and weaknesses for different kinds of analyses. NHANES may be more efficient than EHRs for population-based studies. Firstly, EHRs only contain the data of patients who have paid visits to a hospital, whereas NHANES samples are not limited to patients in hospitals and can represent the U.S. national patient population. Secondly, NHANES data are well structured and readily analyzable, whereas the multi-dimensional data quality problem and largely unstructured data elements of EHRs pose significant impediments for their secondary use. Conversely, EHR data usually contain multiple readings for a variable for a patient, allowing longitudinal analysis of disease progression. Most variables in NHANES have only one reading for each sample. Meanwhile, NHANES may not provide sufficient data for assessing the population representativeness of trials on other medical conditions.

Other works for improving generalizability of trials

To improve the generalizability of clinical trials, Arterburn *et al.* proposed a population-based shared decision-making recruitment approach, which achieved a relatively high ratio of outreach-to-randomized subjects [16]. However, the selection biases with this approach may still remain. Frangakis *et al.* discussed various strategies to calibrate the treatment effects from clinical trials to target populations, including calibration of pre-treatment exclusion criteria and post-treatment measures [17]. However, pre-treatment calibration can only generalize categorical and dichotomous variables but not continuous variables such as age, HbA1c, and BMI. Post-treatment calibration can only generalize components of randomized controlled trials that are not causal effects of treatment [18]. Therefore, these methods may not fully resolve the generalizability issue. We hope that this line of research can help trial designers better balance the tradeoffs between internal validity and external validity during the design phase of a new trial, thereby improving its population representativeness.

Limitations and future work

Quite a few limitations are noteworthy when interpreting this study. NHANES uses self-reported medical condition. We borrowed a method from Dodd *et al.* [11] to distinguish between Type 2 and Type 1 diabetes patients, which, in conjunction with the self-reported data, may have lead to some misclassification of samples.

Even though FDAAA mandates the reporting of basic summary results on registered or approved product to be submitted to ClinicalTrials.gov, only about 20% of T2DM trials between 2003 and 2012 have submitted results to ClinicalTrials.gov. Fewer than half of NIH-funded trials published their results in scientific journals within 30 months of trial completion [19]. Since trials with positive results tend to report them, the convenience sample of study populations we used in this work

may not be a random sample of the enrolled patients in all the T2DM trials. Moreover, the study population plotted in Gaussian distribution may differ from the histogram of the actual patients. Recently, the US Department of Health and Human Services proposed a new regulation to require public sharing of summary data from clinical trials of FDA regulated drugs and devices regardless of their approval status for marketing [20]. As more and more studies publish their results in ClinicalTrials.gov, a uniform reporting mechanism would significantly help aggregate analysis such as the one we performed in this study. Another frequent concern is that clinical trials retrieved by ClinicalTrials.gov's API may have condition-indexing errors, though we checked a random sample of 100 T2DM trials and found that the accuracy of indexing is acceptable for this condition and this study. The accuracy of our natural language processing tool for parsing complex free-text eligibility criteria needs further improvements.

Different eligibility features may have inherent correlations. For example, impaired fasting glucose is positively correlated with age [21]. In the future, we plan to use multiple features simultaneously to assess the population representativeness of trials.

Conclusions

In this study, we identified and visualized statistically significant differences among the T2DM patient population, the target population of T2DM trials, and a convenience sample of the study population extracted from the results of selected T2DM trials in ClinicalTrials.gov. The consistent results with our previous study [4] have demonstrated the feasibility of using NHANES data to assess the population representativeness of clinical trials. We have further validated the results of our previous study using EHR data [4].

Acknowledgments

We thank Dr. Praveen Chandar for the insightful discussion on this work. This study is sponsored by the National Library of Medicine Grant **R01LM009886** (PI: Weng) and the National Center for Advancing Translational Science Award **UL1TR000040** (PI: Ginsberg).

References

- [1] van de Water W, Kiderlen M, Bastiaannet E, Siesling S, Westendorp RG, van de Velde CJ, et al. External validity of a trial comprised of elderly patients with hormone receptor-positive breast cancer. *J Natl Cancer Inst.* 2014;106(4):dju051.
- [2] Blanco C, Olfson M, Goodwin RD, Ogburn E, Liebowitz MR, Nunes EV, et al. Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry.* 2008;69(8):1276-80.
- [3] Hao T, Rusanov A, Boland MR, Weng C. Clustering clinical trials with similar eligibility criteria features. *J Biomed Inform.* 2014;52:112-20.
- [4] Weng C, Li Y, Ryan P, Zhang Y, Gao J, Liu F, et al. A Distribution-based Method for Assessing The Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records. *Applied Clinical Informatics.* 2014;5(2):463-79.
- [5] Rusanov A, Weiskopf NG, Wang S, Weng C. Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. *BMC Med Inform Decis Mak.* 2014;14:51.
- [6] CDC. National Health and Nutrition Examination Survey Data: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention;. Available from: <http://www.cdc.gov/nchs/nhanes.htm>.
- [7] Chen DP, Morgan AA, Butte AJ. Validating pathophysiological models of aging using clinical electronic medical records. *J Biomed Inform.* 2010;43(3):358-64.
- [8] Bays HE, Chapman RH, Grandy S, Group SI. The relationship of body mass index to diabetes mellitus, hypertension and dyslipidaemia: comparison of data from two national surveys. *Int J Clin Pract.* 2007;61(5):737-47.
- [9] He Z, Carini S, Hao T, Sim I, Weng C. A Method for Analyzing Commonalities in Clinical Trial Target Populations. *AMIA Annual Symp Proc.* 2014;In press.
- [10] Description of Laboratory Methodology to Measure Glycohemoglobin in NHANES [October 19, 2014]. Available from: http://www.cdc.gov/nchs/nhanes/2011-2012/GHB_G.htm.
- [11] Dodd AH, Colby MS, Boye KS, Fahlman C, Kim S, Briefel RR. Treatment approach and HbA1c control among US adults with type 2 diabetes: NHANES 1999-2004. *Curr Med Res Opin.* 2009;25(7):1605-13.
- [12] Key Concepts About Weighting in NHANES [October 2014]. Available from: <http://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/Weighting/OverviewKey.htm>.
- [13] CDC. National Health and Nutrition Examination Survey: Analytic Guidelines, 1999-2010 [September 20, 2014]. Available from: http://www.cdc.gov/nchs/data/series/sr_02/sr02_161.pdf.
- [14] Pooled Variance [October 19, 2014]. Available from: http://en.wikipedia.org/wiki/Pooled_variance.
- [15] Cleveland WS, Devlin SJ. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association.* 1988;83:596-610.
- [16] Arterburn D, Flum DR, Westbrook EO, Fuller S, Shea M, Bock SN, et al. A population-based, shared decision-making approach to recruit for a randomized trial of bariatric surgery versus lifestyle for type 2 diabetes. *Surg Obes Relat Dis.* 2013;9(6):837-44.
- [17] Frangakis C. The calibration of treatment effects from clinical trials to target populations. *Clin Trials.* 2009;6(2):136-40.
- [18] Hatfield J, Faunce GJ, Soames Job RF. Avoiding confusion surrounding the phrase "correlation does not imply causation.". *Teaching of Psychology.* 2006;33:49-51.
- [19] Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *BMJ.* 2012;344:d7292.
- [20] Hudson KL, Collins FS. Sharing and Reporting the Results of Clinical Trials. *JAMA.* 2015;313(4):355-6.
- [21] Cowie CC, Rust KF, Byrd-Holt DD, Eberhardt MS, Flegal KM, Engelgau MM, et al. Prevalence of diabetes and impaired fasting glucose in adults in the U.S. population: National Health And Nutrition Examination Survey 1999-2002. *Diabetes Care.* 2006;29(6):1263-8.

Address for correspondence

Dr. Chunhua Weng, 622 W 168th Street, PH-20, New York, NY 10032, USA. Email: cw2384@columbia.edu