Annals of Epidemiology 26 (2016) 198–203

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

A novel method for estimating transgender status using electronic medical records

Douglas Roblin PhD [a,b], Joshua Barzilay MD [b], Dennis Tolsma MPH [b], Brandi Robinson MPH [b], Laura Schild [b], Lee Cromwell MS [b], Hayley Braun [c], Rebecca Nash [c], Joseph Gerth [d], Enid Hunkeler MA, FAHA [e], Virginia P. Quinn PhD [f], Vin Tangpricha MD, PhD [g,h], Michael Goodman MD, MPH [c,*]

[a] School of Public Health, Georgia State University, Atlanta
[b] Center for Clinical and Outcomes Research, Kaiser Permanente Georgia, Atlanta
[c] Rollins School of Public Health, Emory University, Atlanta, GA
[d] Emory College of Arts and Sciences, Atlanta, GA
[e] Division of Research, Kaiser Permanente, Oakland, CA
[f] Kaiser Permanente Southern California, Pasadena
[g] Emory University, School of Medicine, Atlanta, GA
[h] The Atlanta VA Medical Center, Atlanta, GA

ARTICLE INFO

Article history:
Received 20 October 2015
Accepted 6 January 2016
Available online 4 February 2016

Keywords:
Transgender
Electronic medical records
Algorithm
Prevalence

ABSTRACT

Purpose: We describe a novel algorithm for identifying transgender people and determining their male-to-female (MTF) or female-to-male (FTM) identity in electronic medical records of an integrated health system.

Methods: A computer program scanned Kaiser Permanente Georgia electronic medical records from January 2006 through December 2014 for relevant diagnostic codes, and presence of specific keywords (e.g., "transgender" or "transsexual") in clinical notes. Eligibility was verified by review of de-identified text strings containing targeted keywords, and if needed, by an additional in-depth review of records. Once transgender status was confirmed, FTM or MTF identity was assessed using a second program and another round of text string reviews.

Results: Of 813,737 members, 271 were identified as possibly transgender: 137 through keywords only, 25 through diagnostic codes only, and 109 through both codes and keywords. Of these individuals, 185 (68%, 95% confidence interval [CI]: 62%–74%) were confirmed as definitely transgender. The proportions (95% CIs) of definite transgender status among persons identified via keywords, diagnostic codes, and both were 45% (37%–54%), 56% (35%–75%), and 100% (96%–100%). Of the 185 definitely transgender people, 99 (54%, 95% CI: 46%–61%) were MTF, 84 (45%, 95% CI: 38%–53%) were FTM. For two persons, gender identity remained unknown. Prevalence of transgender people (per 100,000 members) was 4.4 (95% CI: 2.6–7.4) in 2006 and 38.7 (95% CI: 32.4–46.2) in 2014.

Conclusions: The proposed method of identifying candidates for transgender health studies is low cost and relatively efficient. It can be applied in other similar health care systems.

Introduction

Transgender people are a heterogeneous group of individuals who transcend normative cultural definitions and categories of sex and gender. Sex is assigned at birth based on primary sexual characteristics [1]. A person's gender refers to one's sense of maleness, femaleness, neither, or both [1,2]. Transgender people are those whose gender identity or expression differs from the sex originally assigned to them at birth [3]. Although self-identification of transgender people may not fit binary definitions [4], a person whose gender identity differs from a male sex assignment at birth is often referred to as male-to-female (MTF) and a person whose gender identity differs from a female sex assignment at birth is often referred to as a female-to-male (FTM) [5]. Transgender people may experience gender dysphoria, which is a diagnostic term that

* Corresponding author. Department of Epidemiology, Emory University School of Public Health, 1518 Clifton Road, NE CNR 3021, Atlanta, GA 30322. Tel.: +1 404-727-2734; fax: +1 404-727-8737.
E-mail address: mgoodm2@emory.edu (M. Goodman).

http://dx.doi.org/10.1016/j.annepidem.2016.01.004
1047-2797/

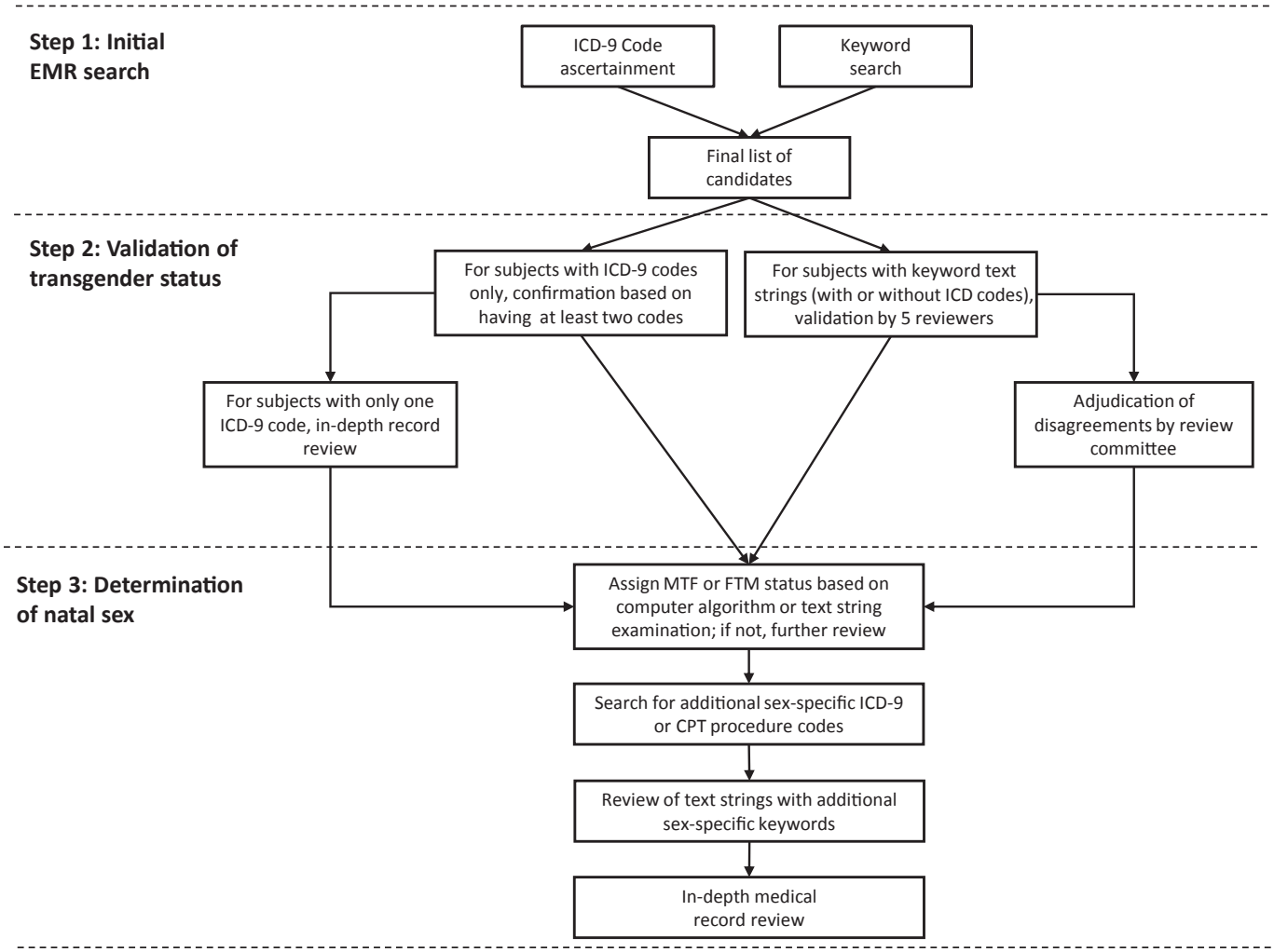

Fig. 1. Flow diagram of assessing transgender status and natal sex among KPGA members.

describes "a discomfort or distress that is caused by a discrepancy between a person's gender identity and that person's sex assigned at birth" [6]. In some cases, gender dysphoria requires gender affirmation, which may include hormonal or surgical treatment, or both [7].

The 2011 Institute of Medicine report on the health of sexual and gender minorities specifically emphasized the need for more information about people who are transgender [8]. Even basic information, such as the proportion of transgender people in the general population, is not known with certainty because most available studies are based on convenience samples without an identifiable population denominator [9].

Electronic medical records (EMR) provide diagnostic codes that offer opportunities for the identification of hard-to-reach subgroups in large well-defined populations. However, for transgender people, reliance on the diagnostic codes for gender dysphoria or related diagnoses is inadequate because many transgender people experience no gender dysphoria and have no need for gender affirmation therapy. In addition, transgender people who are already receiving or wish to receive hormonal or surgical gender affirmation may not have transgender-specific diagnoses documented in the EMR [10]. An alternative option to identify transgender people in medical records could be through natural language processing (NLP) methods. Many NLP methods are complex and require specialized software, substantial expertise, time, and cost to build a reliable computerized algorithm [11,12]. Relatively simple algorithms that use standard software may be more practical.

In this communication, we describe a method for creating and validating a cohort of transgender people using a simple algorithm that combines diagnostic codes and text string-based NLP. We also show how after confirmation of transgender status, the algorithm is useful for identifying each person's MTF or FTM status. We then apply this algorithm to estimate proportion of transgender people among members of an integrated health care system.

Methods

Study setting

This study took place at Kaiser Permanente Georgia (KPGA), an integrated care delivery system that provided health services to approximately 814,000 enrollees between 2006 and 2015. The study was conducted in cooperation with the Emory University School of Public Health. All activities described in this article were reviewed and approved by the Institutional Review Boards of both institutions. KPGA is a member of several research consortia including the Health Care Systems Research Network (formerly known as the HMO Research Network [13]) and the Mental Health Research Network [14]. The 19 health care systems comprising these networks have over 20 million enrollees, use similar EMR