Value of Free-text Comments for Validating Cancer Cases Using Primary-care Data in the United Kingdom

To the Editor:

In the United Kingdom Clinical Practice Research Datalink, free-text comments are unstructured fields for information supplementing coded entries in patients' electronic medical records. Owing to increased governance requirements, free-text comments stopped being available for research in 2016.[1]

Here, we describe the value added by free-text comments in the validation of four common cancer types in primary care data from the Clinical Practice Research Datalink. This validation effort was part of a safety study of antimuscarinic drugs requested by the United States Food and Drug Administration.[2]

For this study, we assembled a cohort of new users of antimuscarinic drugs to treat overactive bladder in years 2004–2012, identified provisional cases of the 10 most common cancers based on cancer diagnosis Read codes, and created electronic medical profiles (all coded entries except study drugs, plus free-text comments) to confirm cancer diagnoses.[3,4]

Reproducibility: The results of the study were generated by RTI Health Solutions (RTI-HS) using data obtained from CPRD. RTI-HS developed proprietary code to perform the analyses on the data. Researchers desiring access to the data would be required to obtain permission from the study sponsor, obtain data use agreement with CPRD, and develop their own code.

This study was funded by Astellas Pharma Global Development, Inc.

The contract provides the research team independent publication rights. A.M., J.F., J.K., B.C., M.R., E.P., L.M., S.P.-G., and A.A. are employees of RTI International, an independent, nonprofit research organization that does work for government agencies and pharmaceutical companies. W.J.A., B.F., and S.d.V. are employees of Astellas Pharma Global Development, the sponsors of this study.

ISSN: 1044-3983/2018/295-0e41
DOI: 10.1097/EDE.0000000000000856

We assessed the relative contribution of free-text comments by comparing the results of the review of medical profiles with and without free-text comments for four cancers: prostate, breast (in women), lung, and bladder. We selected these cancers because they were common and included cancer types that were susceptible to protopathic bias after starting overactive bladder treatment (prostate and bladder cancer).[3] First, one physician reviewed the medical profiles with free-text comments. Second, we removed these comments and a second physician independently reviewed the profiles, following the same criteria and process as previously described.[3] With the review of profiles with free-text comments as the gold standard, we estimated the positive predictive value, negative predictive value, sensitivity, and specificity of the review of profiles without free-text comments, overall and by cancer type. The two reviewers discussed cases with differing status in the reviews with and without free-text comments, taking into account the notes written at the time of the reviews, to understand the reasons for disagreement.

We reviewed electronic patient profiles for 168 provisional cases of bladder, breast, lung, and prostate cancer with and without free-text comments; 143 (85%) were confirmed in the review with free text, and 137 were adjudicated as cases in the review without free text. For the review without free text, the positive predictive value was 0.93 (95% confidence interval [CI] = 0.88, 0.97), negative predictive value was 0.52 (95% CI = 0.34, 0.69), sensitivity was 0.90 (95% CI = 0.84, 0.94), and specificity was 0.64 (95% CI = 0.44, 0.81). Results were similar for individual cancer types (Table).

Of 24 cases for whom results from patient profile review with and without free text were different, 15 (63%) were false-negatives (considered noncases in the review without free text but confirmed as cases in the review with free text); the rest were false-positives. For these 24 discordant cases, free text added relevant information that allowed confirmation as either case or noncase in 15 cases (63%); thus, the absence of free text was responsible for outcome misclassification in 15 potential cases (9% of 168 profiles). In six (25%) of the discordant cases, discrepancies were due to variability in the interpretation of patient profiles and free text by the reviewers (interrater variability).

Although the review without free text classified most cases correctly, free text provided information that was useful to classify case status correctly in 9% of provisional cases. Other researchers using UK primary care data noted that 11% of cancer cases were mentioned only in free-text comments[5] and that free-text comments offered information on cancer staging and treatment that was not always available in coded entries.[6]

Without free-text comments, treatment and staging information may be missed in studies based on UK primary care data, and misclassification of cancer case status will likely increase.

ACKNOWLEDGMENTS

We thank Jennifer Bartsch for her help with programming, Whitney Krueger, Christine Bui, and Alicia Gilsenan for their help managing the project, John Forbes for his editorial help (all from RTI Health Solutions), Kwame Appenteng and Milbhor D'Silva for their input at all stages of the study (both from Astellas), and CPRD research staff for their support to the program.

Andrea V Margulis
amargulis@rti.org

Joan Fortuny
RTI Health Solutions
Barcelona, Spain

James A Kaye
RTI Health Solutions
Waltham, MA

Brian Calingaert

Maria Reynolds
RTI Health Solutions, RTP
Research Triangle Park, NC

Estel Plana
RTI Health Solutions
Barcelona, Spain

Lisa J. McQuay
RTI Health Solutions, RTP
Research Triangle Park, NC

Table. Positive Predictive Value, Negative Predictive Value, Sensitivity, and Specificity of Validation Through Review of Electronic Medical Record Profiles Without Free-Text Comments

	PPV			NPV[a]			Sensitivity			Specificity[a]		
Cancer Type	Num	Den	PPV (95% CI)	Num	Den	NPV (95% CI)	Num	Den	Sensitivity (95% CI)	Num	Den	Specificity (95% CI)
All (N = 168)	128	137	0.93 (0.88, 0.97)	16	31	0.52 (0.34, 0.69)	128	143	0.90 (0.84, 0.94)	16	25	0.64 (0.44, 0.81)
Bladder (n = 36)	27	29	0.93 (0.79, 0.99)	4	7	0.57 (0.22, 0.88)	27	30	0.90 (0.75, 0.97)	4	6	0.67 (0.26, 0.94)
Breast (n = 30)	27	28	0.96 (0.84, 1.00)	2	2	1.0 (0.22, 1.0)	27	27	1.0 (0.89, 1.0)	2	3	0.67 (0.13, 0.98)
Lung (n = 52)	37	40	0.93 (0.81, 0.98)	6	12	0.50 (0.23, 0.77)	37	43	0.86 (0.73, 0.94)	6	9	0.67 (0.33, 0.91)
Prostate (n = 50)	37	40	0.93 (0.81, 0.98)	4	10	0.40 (0.14, 0.71)	37	43	0.86 (0.73, 0.94)	4	7	0.57 (0.22, 0.88)

[a]The low NPV and specificity are explained by the fact that all the profiles had at least one cancer diagnosis code.
Gold standard: results from review of electronic medical records with free-text comments.
CI indicates confidence interval; Den, denominator; NPV, negative predictive value; Num, numerator; PPV, positive predictive value.

Willem Jan Atsma
Billy Franks
Stefan de Vogel
Astellas Pharma B.V.
Leiden, The Netherlands

Susana Perez-Gutthann
Alejandro Arana
RTI Health Solutions
Barcelona, Spain

REFERENCES

1. Independent Scientific Advisory Committee for MHRA database research (ISAC). Summary Minutes of 19 January 2016 Meeting. Available at: https://www.cprd.com/_docs/ISAC%202016%20January%20Final%20Summary%20Minutes.pdf. Accessed 16 November 2017.
2. ENCePP. Post-authorization Safety Program – Validation of the Clinical Practice Research Datalink for the Study of Cardiovascular and Neoplasm Events in Users of Treatments for Overactive Bladder (EUPAS5529). London: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance; 2016. Available at: http://www.encepp.eu/encepp/viewResource.htm?id=11107. Accessed 16 November 2017.
3. Kaye JA, Margulis AV, Fortuny J, et al. Cancer incidence after initiation of antimuscarinic medications for overactive bladder in the United Kingdom: evidence for protopathic bias. *Pharmacotherapy*. 2017;37:673–683.
4. Margulis AV, Fortuny J, Kaye JA, et al. Validation of cancer cases using primary care, cancer registry, and hospitalization data in the UK. *Epidemiology*. 2017 [Epub ahead of print].
5. Boggon R, van Staa TP, Chapman M, Gallagher AM, Hammad TA, Richards MA. Cancer recording and mortality in the General Practice Research Database and linked cancer registries. *Pharmacoepidemiol Drug Saf*. 2013;22:168–175.
6. Cea Soriano L, Soriano-Gabarró M, García Rodríguez LA. Validity and completeness of colorectal cancer diagnoses in a primary care database in the United Kingdom. *Pharmacoepidemiol Drug Saf*. 2016;25:385–391.

OPEN

Long Working Hours and Risk of Venous Thromboembolism

We are unable to provide direct access to the data from the single studies analyzed here. Code is available on request.
The consortium is supported by NordForsk, the UK Medical Research Council (K013351), the Academy of Finland (311492), and a Helsinki Institute of Life Science fellowship for Kivimäki. Funding bodies for participating cohort studies are listed on their Web sites. The study was conducted independently of funding agencies. None of the funding agencies played an active role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the article; and decision to submit the article for publication.
The authors report no conflicts of interest.
SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

*Investigators of the IPD-Work Venous Thromboembolism group include (in alphabetic order): Lars Alfredsson, G. David Batty, Jakob B. Bjorner, Marianne Borritz, Hermann Burr, Nico Dragano, Jane E. Ferrie, Eleonor I. Fransson, Mark Hamer, Katriina Heikkilä, Markus Jokela, Ichiro Kawachi, Anders Knutsson, Mika Kivimäki, Markku Koskenvuo, Ida E. H. Madsen, Martin L. Nielsen, Maria Nordin, Solja T. Nyberg, Tuula Oksanen, Jan H. Pejtersen, Jaana Pentti, Reiner Rugulies, Paula Salo, Archana Singh-Manoux, Sakari B. Suominen, Adam G. Tabak, Töres Theorell, Jussi Vahtera, Marianna Virtanen, Hugo Westerlund, and Peter J. M. Westerholm.

ISSN: 1044-3983/2018/295-0e42
DOI: 10.1097/EDE.0000000000000862

To the Editor:

Venous thromboembolism (VTE) results from a blood clot that forms within a vein.[1] It includes two subtypes: deep-vein thrombosis (a clot in a deep-vein, usually in the leg) and pulmonary embolism (a sudden blockage in a lung artery). Studies of people sleeping in deck chairs in air-raid shelters during the second world war and, more recently, those of passengers on long-haul flights have linked extended periods of sitting to increased VTE risk.[2] It is also the case that psychological stress can unfavorably influence blood coagulation and viscosity, potentially increasing the risk of VTE.[3,4] People working long hours are often characterized by both sedentary behavior and stress, but to our knowledge, no studies are available on the association of this working pattern with VTE. This is therefore the focus of the present analyses.

We drew individual-level data from eight prospective cohort studies participating in the Individual–Participant–Data meta-analysis in Working Populations ("IPD-Work") Consortium.[5] We excluded people not in full-time employment and those with extant disease at study baseline. Working hours and participant characteristics were assessed at baseline (total N = 77,005 to 77,291 depending on the outcome; eAppendix; http://links.lww.com/EDE/B359). All study members were followed up for VTE for a mean of 9.7 years.

As previously,[6–8] we defined ≥55 hours/week as long working hours, with