

Evaluation of Free-Text Comments to Validate Common Cancer Diagnoses in the UK CPRD

Joan Fortuny,¹ James A. Kaye,² Christine L. Bui,³ Alicia W. Gilsenan,³ Jennifer Bartsch,³ Estel Plana,¹ Lisa J. McQuay,³ Brian Calingaert,³ Willem Jan Atsma,⁴ Kwame Appenteng,⁵ Billy Franks,⁴ Stefan de Vogel,⁴ Milbhor D'Silva,⁵ Susana Perez-Gutthann,¹ Alejandro Arana,¹ Andrea V. Margulis¹

¹RTI Health Solutions, Barcelona, Spain; ²RTI Health Solutions, Waltham, MA, United States;

³RTI Health Solutions, Research Triangle Park, NC, United States; ⁴Astellas, Leiden, The Netherlands; ⁵Astellas, Northbrook, IL, United States

CONFLICT OF INTEREST

J. Fortuny, J. Kaye, C. Bui, A. Gilsenan, J. Bartsch E. Plana, L. McQuay, B. Calingaert, S. Perez-Gutthann, A. Arana, and A. Margulis are full-time employees of RTI Health Solutions, which received funding from Astellas Pharma Global Development, Inc. to conduct this study. The contract between RTI Health Solutions and the sponsor includes independent publication rights. RTI conducts work for government, public, and private organizations, including pharmaceutical companies. W.J. Atsma, K. Appenteng, B. Franks, S. de Vogel, and M. D'Silva are employees of Astellas Pharma Global Development, Inc.

BACKGROUND

- Some primary care databases (e.g., the Clinical Practice Research Datalink [CPRD]) include free-text comments, which reflect physicians' thinking without constraint to coded entries.
- Free text has been used in database studies to help validate outcomes.¹
- Starting in April 2016, free text is no longer available for research in the CPRD, owing to transparency and governance concerns.
- The loss of this source of patient data may potentially affect the validity of clinical information obtained from the CPRD.
- In the context of a larger postapproval safety study² requested by the health authorities for mirabegron (a beta-3 adrenergic agonist indicated for the treatment of symptomatic overactive bladder [OAB]), the risk of cancer in users of antimuscarinic OAB medication was evaluated in the CPRD using free text to assist in the validation of cancer outcomes. As part of a validation program, we assessed the relative contribution of free text in the validation of four common cancers.

OBJECTIVE

- To evaluate the relative contribution of free-text comments in the confirmation of incident cases of prostate, breast, lung, and bladder cancer in the CPRD, using Hospital Episode Statistics (HES) and cancer registry linkages.

METHODS

Identification of Study Population

- Step 1:** We developed an electronic screening algorithm to identify potential cancer cases in the CPRD GOLD and retained patients with diagnostic codes for prostate, breast, lung, and bladder cancer.
- Step 2:** We excluded patients who had coded diagnoses of the cancers of interest in HES or the cancer registry; these patients were considered automatically confirmed.
- Step 3:** We excluded patients who had clear evidence of the cancers of interest and no evidence for other cancers; these patients were considered confirmed.
- Step 4:** Among the remaining patients (i.e., those with unclear diagnoses for the cancers of interest based on Read codes in the CPRD GOLD), we retained those who had free text available in their medical records.

Creation of Electronic Medical Records

- For the same group of patients, we created two sets of electronic medical record profiles (including information on prescriptions, diagnoses, procedures, laboratory tests, referrals, and clinical information), one that included the available free text and one without the free text.

Identification and Confirmation of Cancer Cases (Figure 1)

- To decrease interrater variability, two study physicians underwent training on how to review profiles.
- The reviews were conducted on the original set of profiles used for the validation phase of the larger postapproval safety study.^{2,5}
- Reviewer 1 reviewed profiles without free text (i.e., where free text had been redacted), relying only on Read-coded data in the CPRD GOLD.
- Reviewer 2 reviewed profiles with free text (i.e., original profiles used in the overarching study), relying on Read-coded data in CPRD GOLD and any associated free-text annotations.
- Each reviewer independently determined cancer type (e.g., breast, bladder) and confirmation status (i.e., not confirmed or confirmed).
 - Not confirmed:** cancer diagnosed date was outside the study period (e.g., before exposure to the antimuscarinic medication that determined cohort entry), an incident cancer was ruled out, or information was insufficient to confirm the diagnosis.
 - Confirmed in the CPRD GOLD:** Read codes indicated cancer treatment, repeated use of cancer diagnostic codes, or the use of a subsequent "cancer care review" code (used by the National Institute for Health and Care Excellence as an indicator in the Quality and Outcomes Framework program; reflects that the general practitioner has reviewed the follow-up of a cancer patient).⁶
 - Confirmed via free text:** further details on cancer diagnoses, as recorded by the general practitioner, provided substantiating information.
- Following the independent review of each case by both physicians described above, the physicians jointly reviewed cases with discordant results to identify possible reasons for the discrepancy. Free text was available for this evaluation of discrepancies.

RESULTS

- The electronic screening algorithm identified 168 potential cases of the four cancers of interest in the CPRD GOLD that were not confirmed with the evidence in the CPRD GOLD, HES, or cancer registry data.
- Patient profile review of the 168 potential cases by two physicians, stratified by confirmation status after review of free text, resulted in the following:
 - Review without free text (Reviewer 1):
 - 137 cases (128 true cases confirmed with free text [90% of the 143 cases confirmed with free text]; 9 incorrect classifications [7% of the 137 classified as cases by review without free text])
 - 31 noncases
 - Review with free text (Reviewer 2):
 - 143 confirmed cases (considered the gold-standard for the present study)
 - 25 unconfirmed cases
- There were 24 patients (14% of the 168 total potential cases) in whom the results from patient profile reviews (with and without free text) were discrepant:
 - 15 patients (9% of the 168 total potential cases) classified as noncases by the review without free text were confirmed as cancer cases by the information added by free text (i.e., were false negatives on review without free text).
 - 12 patients had additional evidence in free text that allowed confirmation.
 - 3 patient profiles were interpreted differently by the two reviewers (interrater variability), even after both reviewers had access to the free text.
 - 9 patients (5% of the 168 total potential cases) classified as cases by the review without free text were reclassified as noncases when free text was reviewed (i.e., were false positives on review without free text):
 - 3 patients had other diagnoses or insufficient evidence.
 - 3 patients had cancer diagnosis dates reset before study initiation.
 - 3 patient profiles were interpreted differently by the two reviewers (interrater variability), even after both reviewers had access to the free text.
- Overall, interrater variability was 4% (6 of 168 patients).
- Cancer type (i.e., prostate, breast, lung, bladder) determined by review without free text matched in 142 of 143 cases confirmed with free-text review (data not shown).
- Figures 2 and 3 show two examples of the kind of information present in free text that helped confirm cancer cases.
 - Example profiles are shown for illustrative purpose only; while they are based on actual profiles from the CPRD, they have been modified so as not to represent any identifiable person (but relevant free text was not altered).
- Figure 4 shows an example of an instance where the confirmed cancer status was lost for study purposes when free text was used.
 - The example profile, while based on an actual profile from the CPRD, has been modified so as not to represent any identifiable person (but relevant free text was not altered).
- Table 1 presents the overall performance metrics for case classification in the review without free text, which were calculated using the review with free text as the gold standard.

Figure 1. Case Classification Process and Confirmation Status of Cancer According to Free-Text Availability

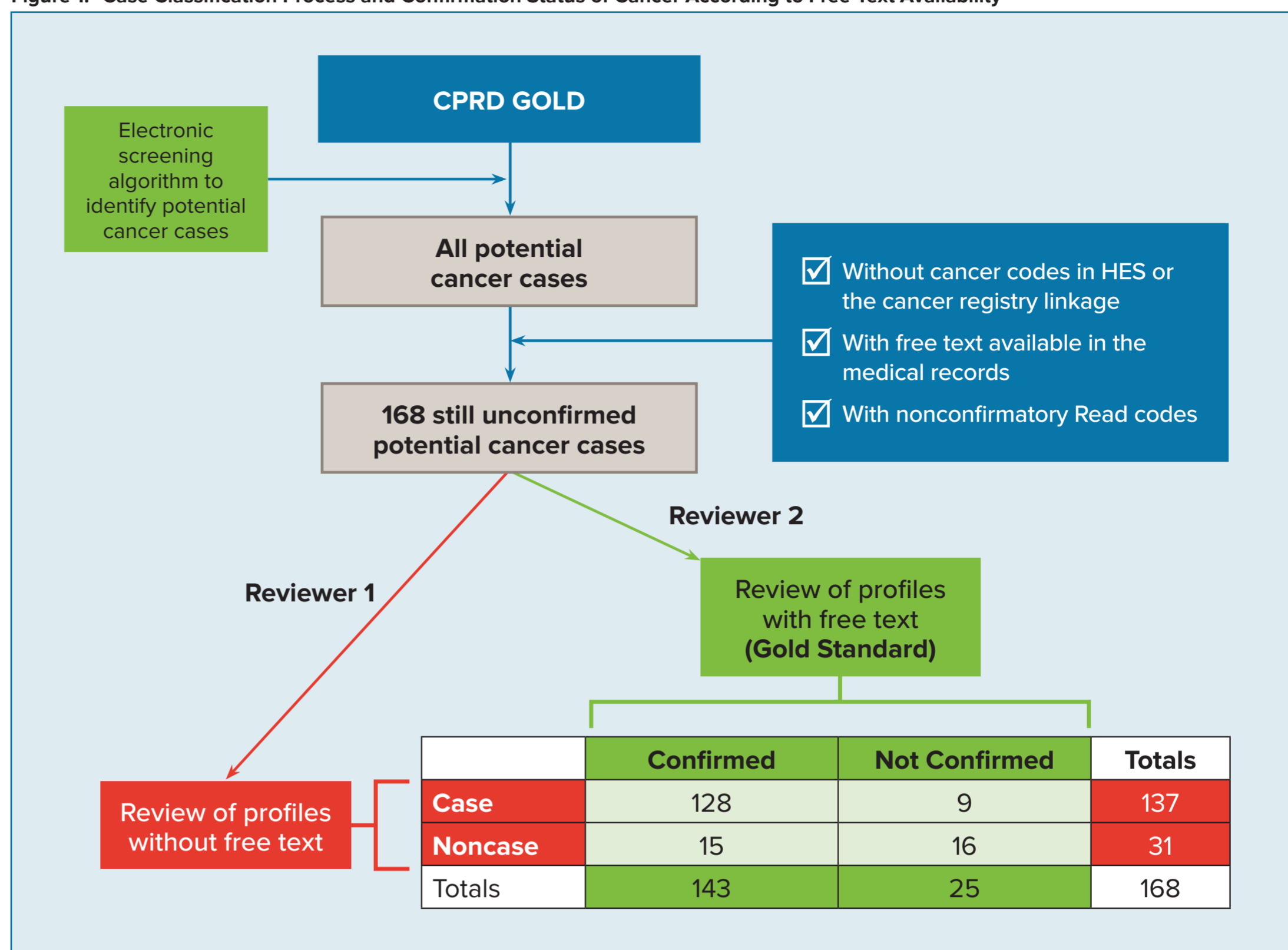


Table 1. Positive Predictive Value, Negative Predictive Value, Sensitivity, and Specificity of Validation (With 95% Confidence Intervals) Through Review of Electronic Medical Record Profiles Without Free-Text Comments

Cancer Type	Positive Predictive Value	Negative Predictive Value	Sensitivity	Specificity
All (N = 168)	0.93 (0.88-0.97)	0.52 (0.34-0.69)	0.90 (0.84-0.94)	0.64 (0.44-0.81)
Bladder (n = 36)	0.93 (0.79-0.99)	0.57 (0.22-0.88)	0.90 (0.75-0.97)	0.67 (0.26-0.94)
Breast (n = 30)	0.96 (0.84-1.00)	1.00 (0.22-1.00)	1.00 (0.89-1.00)	0.67 (0.13-0.98)
Lung (n = 52)	0.93 (0.81-0.98)	0.50 (0.23-0.77)	0.86 (0.73-0.94)	0.67 (0.33-0.91)
Prostate (n = 50)	0.93 (0.81-0.98)	0.40 (0.14-0.71)	0.86 (0.73-0.94)	0.57 (0.22-0.88)

Note: review of electronic medical records with free text is the gold standard.

Figure 2. Example of a Modified Patient Profile Where Free Text Provided Details on a Prostate Cancer That Was Only Mentioned in an Isolated Read Code in the CPRD GOLD

Date	Read Code	Read Code Description	Free Text
2/15/2014	B46..00	Malignant neoplasm of prostate	CA PROSTATE , GLEASON GUG/L 5% OF, TISSUE, PSA AT DIAGNOSIS 71
2/15/2014	9D1..00	MED3 - doctor's statement	C13 PROSTATIC CA/ONGOING LOIN PAIN
9/15/2014	9N36.11	Letter from consultant	DATE ****/****/2014. HOSP ----- UROLOGY ***** CLINIC ----- UROLOGY CONSULTANT----- UROLOGY START MEDICATION-----TAMSULOSIN 400MCG MANE FU

Figure 3. Example of a Modified Patient Profile Where Free Text Made Reference to a Visit to a Cancer Center and Palliative Care Meeting in a Patient With an Otherwise Isolated Bladder Cancer Read Code in the CPRD GOLD

Date	Read Code	Read Code Description	Free Text
10/28/2010	8HJJ.00	Self-referral to accident and emergency department	
8/11/2010	B49..00	Malignant neoplasm of urinary bladder	
8/11/2010	9N11.00	Seen in urology clinic	CLINICAL LETTER *** HOSPITAL UROLOGY
11/11/2010	9N36.11	Letter from consultant	CLINICAL LETTER *** HOSPITAL CANCER CENTRE
11/20/2010	9N1yE00	Seen in physiotherapy department	CLINICAL LETTER *** HOSPITAL PALLIATIVE CARE MULTIDISCIPLINARY MEETING

Figure 4. Example of a Modified Patient Profile Where Free Text Suggested That the Cancer May Have Been Secondary and Therefore Considered Not Confirmed for Study Purposes

Date	Read Code	Read Code Description	Free Text
12/22/2008	B22z.11	Lung cancer	...PROBABLY SECONDARY
12/22/2008	5201	Movicol oral powder 13.8g sachets lemon & lime (Norgine Pharmaceuticals Ltd)	Sodium/Chloride/Potassium/Bicarbonate/Macrogol '3350'
10/1/2009	9OG..00	Geriatric screen admin.	- INVITATION LETTER

DISCUSSION

- The assessment of the relative contribution of free-text comments in the confirmation of cancer cases in the CPRD was conducted in a population of potential cases of prostate, breast, lung, and bladder cancer that was not already confirmed using available HES or cancer registry data (step 2 listed in Methods) nor based on Read codes in patient medical histories. Therefore, the patients included represent a sample of cases for which classification was less clear.
- The absence of free text had a limited effect on the capacity to correctly identify cases from the four common types of cancers studied (93% positive predictive value for the review without free text).
- The review with free text confirmed 12% (15/128) more true cases than the review without free text (an increase from 128 to 143 patients).
- Findings among the 24 discrepant patients were as follows:
 - In 15 patients (63% of those with discrepant reviews), free text added relevant information that led to the reclassification of the patient to confirmed status.
 - In 3 patients (13%), the dates of the cancer diagnosis were reset before the antimuscarinic OAB medication was initiated (effectively making the cases not confirmed for study purposes) on the basis of findings in free text.
 - In 6 patients (25%), discrepancies were due to variability in the interpretation of free text and patient profiles by the reviewers (interrater variability).
- Lack of free text led to moderate misclassification of the status of patients as cases:
 - Overall, 9 of the 137 (7%) patients considered to be cases by the review without free text were classified as such incorrectly. However, in only 3 patients was the misclassification due to a patient not having cancer (3 others were due to changes in date of diagnosis and 3 to interrater variability in the review of the profile).
 - A high proportion (48%) of patients not considered to be cases by the review without free text (15 of 31 patients) were incorrectly classified and were, in fact, confirmed as cases by the review with free text.

CONCLUSIONS

- Review without free text correctly classified most patients. Free text did not add information on cancer type.
- However, about half (15 of 31) of patients not considered cases in the review without free text actually had a cancer of interest, and more than one third (9 of 25) of patients not confirmed by free text were falsely considered cases in the review without free text.
- Interrater variability was low at 4% overall.
- In our study, misclassification of case status (confirmed vs. not confirmed) increased without availability of free text.

REFERENCES

- 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;22:168-75.
- European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (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. 2016. <http://www.encepp.eu/encepp/viewResource.htm?id=1107>. Accessed May 31, 2016.
- Fortuny J, Kaye JA, Margulis A, Plana E, Calingaert B, Perez-Gutthann S, et al. Validity of cancer diagnoses in general practitioner medical records [Abstract 899]. *Pharmacoepidemiol Drug Saf*. 2015;24 Suppl 11-587.
- Kaye JA, Margulis A, Plana E, Calingaert B, Perez-Gutthann S, Arana A. Cancer rates over time after initiation of overactive bladder drugs [Abstract 52]. *Pharmacoepidemiol Drug Saf*. 2015;24 Suppl 11-587.
- Margulis A, McQuay L, Perez-Gutthann S, Kaye JA, Arana A. Use of overactive bladder medications in the adult population of the UK: a cohort study in the Clinical Practice Research Datalink [Abstract 215]. *Pharmacoepidemiol Drug Saf*. 2015;24 Suppl 11-587.
- National Institute for Health and Care Excellence. Quality and outcomes framework (QOF) indicator development programme: indicator guidance. August 2012. <https://www.nice.org.uk/Media/Default/standards-and-indicators/qof%20indicator%20key%20documents/NM62-QOF-Indicator-Guidance-Cancer.pdf>. Accessed August 18, 2015.

CONTACT INFORMATION

Joan Fortuny, MD, PhD
Sr. Research Epidemiologist
RTI Health Solutions
Trav. Gràcia 56 Àtic 1
08006 Barcelona, Spain
Phone: +34.933.622.814
E-mail: jfortuny@rti.org

ABSTRACTS FROM THIS PROGRAM ALSO PRESENTED IN THIS CONFERENCE

- Arana A, et al. **Do individual antimuscarinic drugs to treat overactive bladder have different cardiovascular risks? A UK CPRD cohort study.** Abstract #920. Poster Session C: Safety & Effectiveness - GU & Hormones, Sunday, 28 August 2016, 8:00 AM-1:45 PM.
- Hallas J, et al. **Incidence of cardiovascular events in new users of overactive bladder medications in Denmark.** Abstract #848. Oral presentation in session CV Adverse Events: Affairs of the Heart, Sunday, 28 August 2016, 3:15 PM-4:45 PM.
- Hallas J, et al. **Elevated bladder and prostate cancer rates following initiation of OAB medication: findings from the Danish registries, 2008-2012.** Abstract #918. Poster Session C: Safety & Effectiveness - GU & Hormones, Sunday, 28 August 2016, 8:00 AM-1:45 PM.
- Linder M, et al. **Cancer risk in users of antimuscarinic drugs for overactive bladder: a cohort study in the Swedish national registers.** Abstract #919. Poster Session C: Safety & Effectiveness - GU & Hormones, Sunday, 28 August 2016, 8:00 AM-1:45 PM.
- Linder M, et al. **Cardiovascular risk in users of antimuscarinic drugs for overactive bladder: a cohort study in the Swedish national registers.** Abstract #849. Oral presentation in session CV Adverse Events: Affairs of the Heart, Sunday, 28 August 2016, 3:15 PM-4:45 PM.
- Margulis AV, et al. **Patterns of use of antimuscarinic drugs to treat overactive bladder in Denmark, Sweden, and the United Kingdom.** Abstract #1126. Poster Session C: DUR - Trends in GU and Hormones, Sunday, 28 August 2016, 8:00 AM-1:45 PM.
- Margulis AV, et al. **Validation of cardiovascular events and covariates in CPRD GOLD using questionnaires to general practitioners.** Abstract #437. Oral presentation in session Identification and Validation of Outcomes, Saturday, 27 August 2016, 8:00 AM-9:30 AM.