Validity of Cancer Diagnoses in General Practitioner Medical Records

Joan Fortuny,[1] James A Kaye,[2] Andrea V Margulis,[1] Estel Plana,[1] Brian Calingaert,[3] Susana Perez-Gutthann,[1] Alejandro Arana[1]

[1] RTI Health Solutions, Barcelona, Spain; [2] RTI Health Solutions, Waltham, MA, United States; [3] RTI Health Solutions, Research Triangle Park, NC, United States

BACKGROUND

- An observational study was conducted to prospectively evaluate the incidence of new cancer events (excluding non–melanoma skin cancer) in patients using pharmacological treatments for overactive bladder (OAB).
- To obtain reliable results from the planned study, all the investigated cancer outcomes were validated.
- The study has been registered in the EU PAS Register.

OBJECTIVES

- To investigate the validity of the diagnoses of several common cancers in a cohort of subjects with OAB symptoms treated with antimuscarinic drugs.
- To evaluate the relative contributions of different data sources in the Clinical Practice Research Datalink (CPRD) in the identification of cancer cases during the period from 2004 to 2012.

METHODS

Data Sources: The CPRD

- The CPRD contains the information recorded by general practitioners (GPs) as part of their routine clinical practice and covers approximately 8% of the United Kingdom population.
  - Patients are representative of the whole United Kingdom population in terms of age and sex.
  - Core data include information on diagnoses, symptoms, referrals, tests ordered, test results, prescriptions issued, and additional clinical information.
- Currently, about 65% of the practices contributing to the CPRD have consented to link GP medical records (CPRD GOLD) to other health care data sources via the patient's National Health Service number, sex, date of birth, and postal code.
  - These linked data sources include hospitalization records (Hospital Episode Statistics [HES] data), national mortality data (Office for National Statistics [ONS]), and cancer registry data (National Cancer Data Repository [NCDR]) (Figure 1).
- For all practices, data were obtained from CPRD GOLD for the entire study period.
- For the practices that permit linkage, data were collected as follows:
  - HES data: April 1, 1997 through March 2012
  - Cancer registry (NCDR) data: January 1, 1985 through December 2010

Study Subjects and Follow-up

- A study cohort of new users of OAB drugs was selected with the following criteria:
  - Have at least 12 months of continuous enrollment in the database before cohort entry
  - Have a first recorded prescription for oxybutynin, tolterodine, darifenacin, solifenacin, trospium, or fesoterodine (in decreasing order by frequency) during the study period, without a prescription for the same medication in the 12 months before cohort entry date
  - Be aged 18 years or older at the time of the prescription granting study entry
  - Do not have a diagnosis of cancer (other than non–melanoma skin cancer) prior to cohort entry
  - Do not have a diagnosis of human immunodeficiency virus (HIV) infection prior to cohort entry
- The study period was January 31, 2004 through December 31, 2012.
  - Follow-up started at the index prescription (cohort entry date) and ended at the earliest of the following: end of the study period, disenrollment, diagnosis of HIV or any cancer (except non–melanoma skin cancer), or death.

Cancer Outcomes of Interest

- The present study focuses on the 10 most commonly occurring malignancies in westernized societies, presented below from most to least common.
  - Female breast cancer
  - Lung cancer
  - Prostate cancer
  - Colorectal cancer
  - Melanoma
  - Non-Hodgkin's lymphoma
  - Bladder cancer
  - Renal cancer
  - Pancreatic cancer
  - Uterine cancer

Validation Process

Linked Practices (Figure 2)

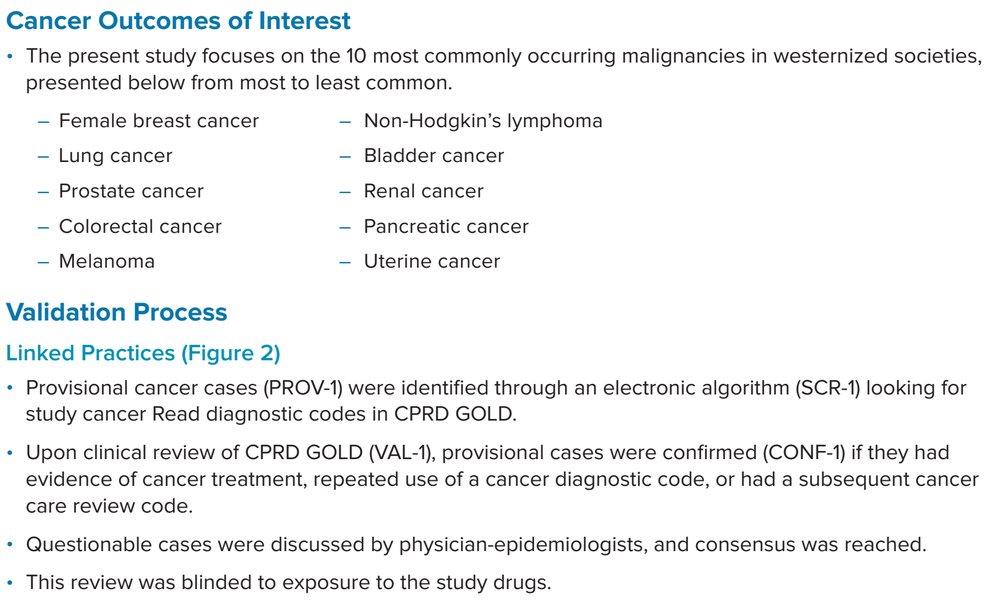
- Provisional cancer cases (PROV-1) were identified through an electronic algorithm (SCR-1) looking for study cancer Read diagnostic codes in CPRD GOLD.
- Upon clinical review of CPRD GOLD (VAL-1), provisional cases were confirmed (CONF-1) if they had evidence of cancer treatment, repeated use of a cancer diagnostic code, or had a subsequent cancer care review code.
- Questionable cases were discussed by physician-epidemiologists, and consensus was reached.
- This review was blinded to exposure to the study drugs.
- After the screening process, subjects considered noncases (Not Case) during review of CPRD GOLD (VAL-1), or those who remained provisional (PROV-1) after review of CPRD GOLD, could become confirmed cases if they were also identified in the cancer registry (SCR-2) or HES (SCR-3) linked data (CONF-2, CONF-3, and CONF-4) by study cancer International Classification of Diseases codes.
- Additional cases could be identified in these data sources without having been identified through screening of CPRD GOLD.

Nonlinked Practices

- For nonlinked practices, potential cases were initially identified with the same screening of CPRD GOLD as was used in the linked practices, but the validation process to confirm cases involved only the clinical review of CPRD GOLD and discussion of questionable cases (CONF-1).

RESULTS

Figure 2. Review and Validation Process in Linked Practices



[a] A total of 241 patients were excluded prior to SCR-1, because they were identified via cancer registry data as having cancer prior to cohort entry. Of these patients, 19 were classified as PROV-1, per SCR-1, and 8 were identified as having prior cancer, per VAL-1.

[b] One case classified as breast cancer per SCR-1 was reclassified to non-Hodgkin's lymphoma per VAL-1, and another screened as bladder plus prostate cancer was determined to be only bladder cancer.

[c] An earlier study cancer of different type was identified per SCR-2 or SCR-3.

[d] An earlier study cancer of different type was identified per SCR-3.

Figure 1. Linkage of Data Sources in CPRD

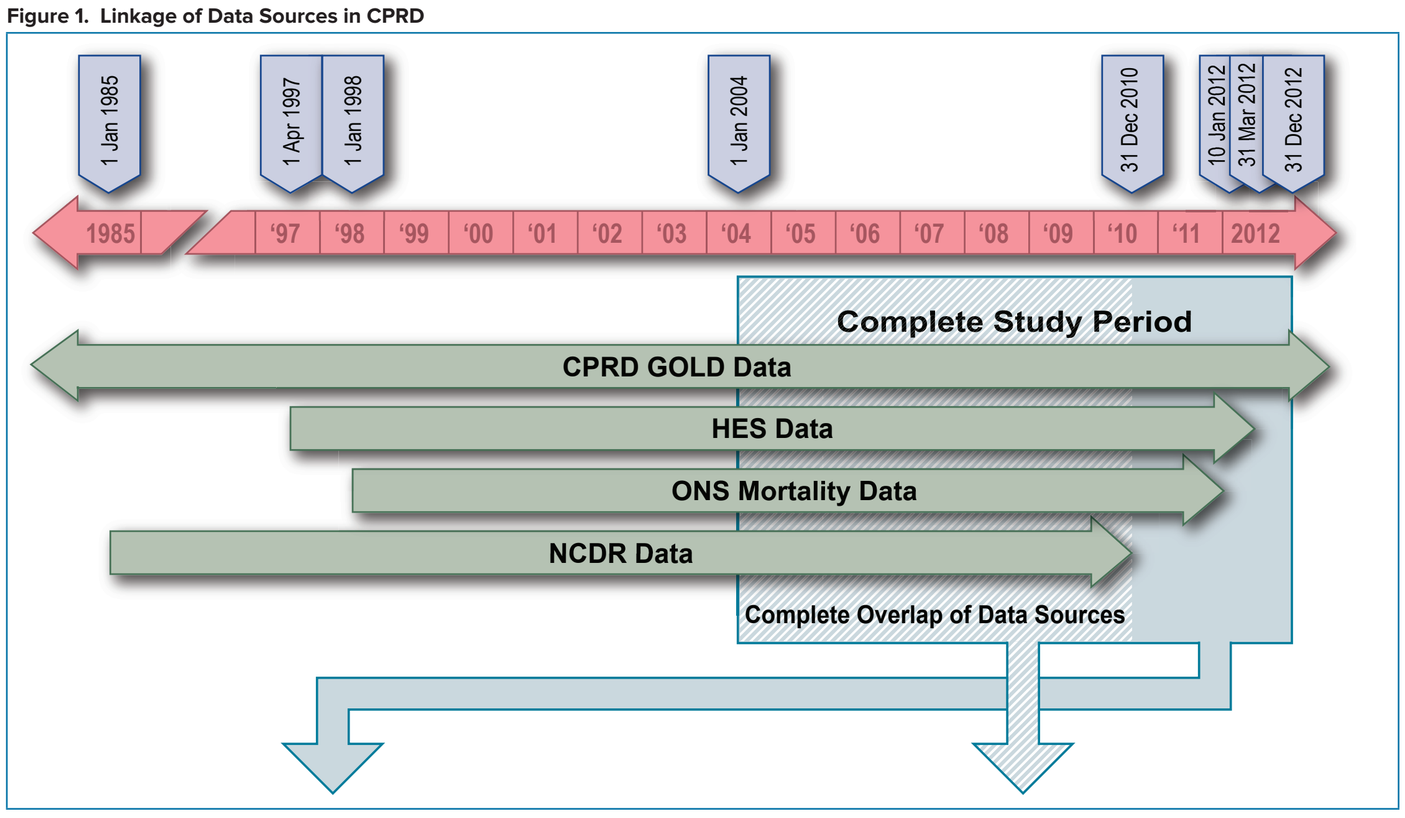


Figure 3. Origin of Cancer Data by Source: Linked Practices, Complete Study Period (2004-2012)

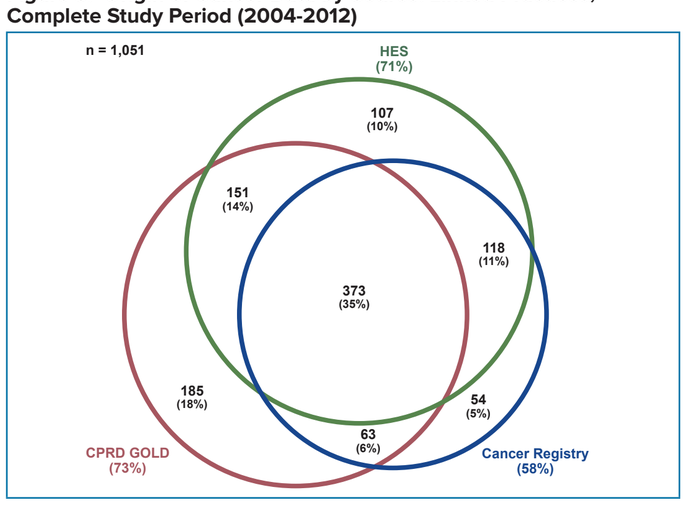
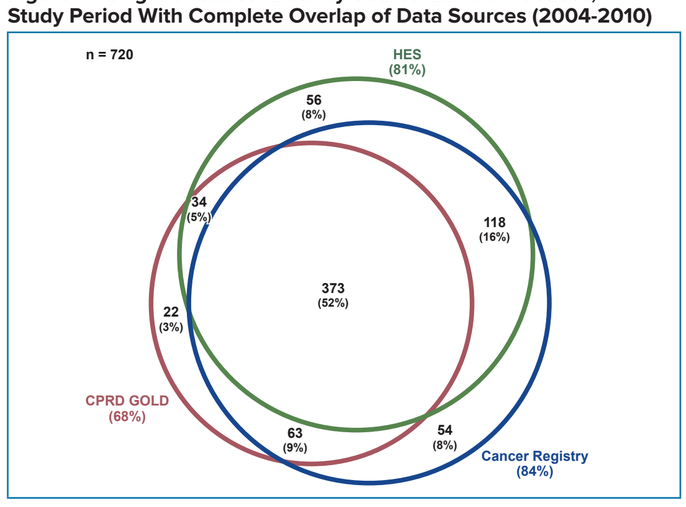
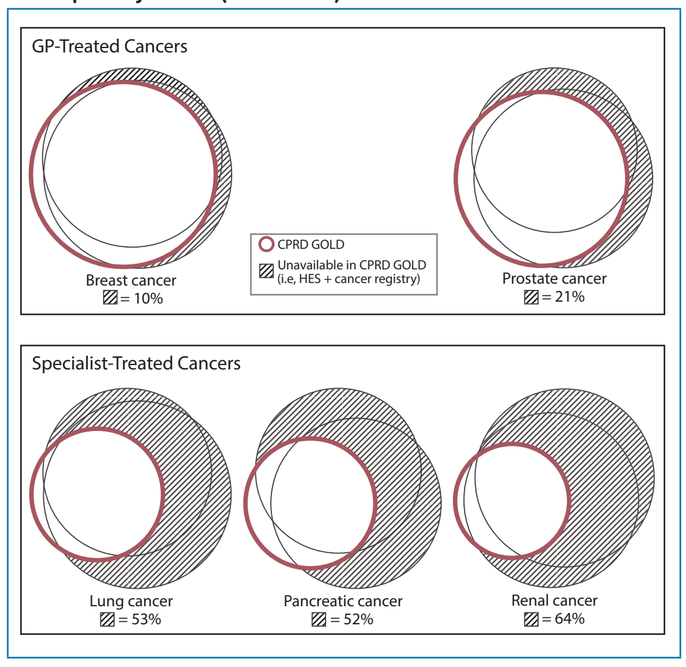


Figure 4. Origin of Cancer Data by Source: Linked Practices, Study Period With Complete Overlap of Data Sources (2004-2010)



- After excluding patients identified by the HES or the cancer registry as having a cancer prior to cohort entry, the electronic search identified 1,457 provisional cancer cases (PROV-1) in CPRD GOLD, 796 from linked practices and 661 from nonlinked practices.
  - Clinical review of CPRD GOLD and subsequent discussion of doubtful cases (VAL-1) confirmed 771 (97%) cases (CONF-1) in linked practices and 616 (93%) (CONF-1) in nonlinked practices.
- When considering the complete study period (2004-2012), 1,051 cases were confirmed in linked practices using all available data sources (i.e., adding cancer registry and/or HES data to the information from CPRD GOLD (CONF-1, CONF-2, CONF-3, and CONF-4).
  - Of these, 73% were identified in CPRD GOLD, 58% in cancer registry data, and 71% in HES; 18% were in CPRD GOLD only, 10% in HES only, and 5% in cancer registry only; 32% were in two sources, and 35% were in all three sources.
  - Figure 3 shows the source of cancer data in linked practices for the complete study period.
- When considering the study years with full overlap of the three sources of data (2004-2010), 720 cases were confirmed in linked practices using all available data sources (CONF-1, CONF-2, CONF-3, and CONF-4).
  - Of these, 68% were identified in CPRD GOLD, 84% in cancer registry data, and 81% in HES; 3% were in CPRD GOLD only, 8% in cancer registry data only, and 8% in HES only; 30% were in two sources, and 52% were in all three sources.
  - Figure 4 shows the source of cancer data in linked practices for the study period with complete overlap of data from the three sources.
- When using the complete study period (2004-2012), 18% of confirmed cases were identifiable only through CPRD GOLD (i.e., no cancer diagnosis data were available in the cancer registry or HES) (Figure 1); however, this figure decreased to 3% when the study period was limited to time with overlap of the three data sources (2004-2010) (Figure 4).
- During the period of overlap, 32% of the confirmed cancer cases would not have been identified using only CPRD GOLD.
  - However, this percentage seemed to vary according to whether GPs typically provide specific medical therapies for each type of cancer. In cancers for which GPs routinely prescribe hormonal treatments, 10% of all breast cancers and 21% of all prostate cancer cases were not identifiable in CPRD GOLD. The corresponding figures for cancers usually treated with parenteral anticancer medications by specialists in referral clinics, such as lung, pancreatic, and renal cancers, were 53%, 52%, and 64%, respectively (Figure 5).

Figure 5. Selected Cancers by Main Treating Physician: Percentage of Cases Not Identifiable in GOLD: Linked Practices, Complete Overlap Study Period (2004-2010)



CONCLUSIONS

- Nearly all cancers recorded in CPRD GOLD (similarly for linked and nonlinked practices) are confirmed by individual profile review or with data from other sources.
- A substantial proportion of cancers will be missed if cancer registry and HES data are not used or are not available. The percentage of cancers absent in CPRD GOLD is higher for cancers in which specific anticancer treatment is not typically prescribed by GPs.
- Different update lags of the three sources of diagnostic data in the CPRD (CPRD GOLD, cancer registry, and HES) affect the relative contribution of each data source to the overall number of cases detected. If the study is limited to a period with complete overlap, the three data sources share a larger proportion of cancer cases.

CONFLICT OF INTEREST STATE

The authors 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.

Abstracts From This Program Also Presented in This Conference

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. Poster Session A: Drug Utilization Research/Health Services Research, Monday, 24 Aug, 8:00 AM-6:00 PM.

Kaye JA, Margulis AV, Plana E, Calingaert B, Perez-Gutthann S, Arana S. Cancer rates over time after initiation of overactive bladder drugs. Abstract #52. Oral presentation in session Those That Can Heal Can Harm; Those that Can Cure Can Kill, Monday, 24 Aug, 1:30 PM-3:00 PM.

Mortimer KM, Ezzy SM, Jessup JT, Gately RV, Seeger JD. Medical record validation of algorithms for acute myocardial infarction (AMI) within a United States (US) administrative claims database. Abstract #1007, Poster Session C: Methods, Wednesday, 26 Aug, 8:00 AM-1:45 PM

Mortimer KM, Ezzy SM, Jessup JT, Gately RV, Seeger JD. Medical record validation of algorithms for ten types of cancer within a United States (US) administrative claims database. Abstract #1016, Poster Session C: Methods, Wednesday, 26 Aug, 8:00 AM-1:45 PM

Mortimer KM, Ezzy SM, Jessup JT, Gately RV, Seeger JD. Revisions to published algorithms for stroke within a United States (US) administrative claims database. Abstract #1009, Poster Session C: Methods, Wednesday, 26 Aug, 8:00 AM-1:45 PM

CONTACT INFORMATION

Joan Fortuny, MD PhD
Senior Research Epidemiologist

RTI Health Solutions
Trav. Gràcia 56 Àtic 1
08006 Barcelona, Spain

Phone: +34.93.362.2814
Fax: +34.93.414.2610
E-mail: jfortuny@rti.org