NON-INTERVENTIONAL (NI) FINAL STUDY REPORT

PASS information

Title	Post Authorization Safety Study (PASS) of Conjugated Estrogens/Bazedoxifene (CE/BZA) in the US
Protocol number	B2311060
Version identifier of the final study report	1.0
Date	26 March 2021
EU Post Authorization Study (PAS) register number	EUPAS11599
Active substance	Conjugated estrogens/bazedoxifene (CE/BZA)
Medicinal product	DUAVEE® 0.45 mg/20 mg tablets
Product reference	**EU MA number: EU/1/14/960/001** (EU marketing authorization granted 16 December 2014)
Procedure number	**EMEA/H/C/002314/MEA 002**
Marketing Authorization Holder (MAH)	Pfizer Europe MA EEIG
Joint PASS	No
Research question and objectives	The overall aim of this PASS was to monitor the safety profile of Duavee (CE/BZA) in comparison to estrogen and progestin combination hormone therapy (E+P HT).

	Towards this end, the following primary and secondary objectives were completed: **<u>Primary Objective:</u>** To estimate the incidence and compare the risks of endometrial hyperplasia and endometrial cancer among postmenopausal women initiating CE/BZA and postmenopausal women initiating E+P HT during the first five years of CE/BZA availability in the US (2014-2019). **<u>Secondary Objectives:</u>** To estimate the incidence and compare the risks of the following safety outcomes among postmenopausal women initiating CE/BZA and postmenopausal women initiating E+P HT: • Venous thromboembolism (VTE) • Myocardial infarction (MI) • Stroke (including transient ischemic attack [TIA]) • Breast cancer • Ovarian cancer • Thyroid cancer • Renal cancer and adenoma • Gastrointestinal cancer • Any cancer (any malignant neoplasm, including cancers listed above and excluding non-melanoma skin cancer [NMSC])
Country(-ies) of study	United States
Author(s)	Renu Garg, PhD, MPH Sarah Hoffman, PhD, MS, MPH

Marketing Authorization Holder(s)

Marketing Authorization Holder(s)	Pfizer Europe MA EEIG Boulevard de la Plaine 17 1050 Bruxelles Belgium
MAH contact person	Renu Garg, PhD, MPH Safety Surveillance Research Scientist Email: Renu.Garg@pfizer.com

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Annex 1. List of stand-alone documents

Appendix 1. SIGNATURES

Appendix 2. PROTOCOL

Appendix 3. INVESTIGATORS AND CORRESPONDING INDEPENDENT ETHICS COMMITTEES (IECs) OR INSTITUTIONAL REVIEW BOARDS (IRBs)

Refer to section 3 Investigators and section 5 Milestones

Appendix 4. STATISTICAL ANALYSIS PLAN

Final Statistical Analysis Plan (21 February 2020)

Appendix A: Code Lists

Appendix B: Table Shells

Appendix C: Figures Illustrating Study Design

Appendix D: Drug Codes

Appendix 5. SAMPLE CASE REPORT FORM (CRF) / DATA COLLECTION TOOL (DCT)

Not applicable.

Appendix 6. SAMPLE STANDARD SUBJECT INFORMATION SHEET AND INFORMED CONSENT DOCUMENT (ICD)

Not applicable.

Appendix 7. LIST OF SUBJECT DATA LISTINGS

Not applicable.

Appendix 8. ADDITIONAL DOCUMENTS

Appendix 8a. Healthagen Site-Specific Results

Appendix 8b. HealthCore Site-Specific Results

Appendix 8c. MarketScan CCAE & Medicare Supplemental Site-Specific Results

Appendix 8d. MarketScan Medicaid Site-Specific Results

Appendix 8e. Optum Site-Specific Results

1. ABSTRACT (STAND-ALONE DOCUMENT)

See accompanying abstract.

2. LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	Adverse event
AR	Assessment report
ASD	Absolute standardized difference
ASO	Administrative services only
CCAE	Claims and commercial encounters
CE/BZA	Conjugated estrogens/bazedoxifene
CI	Confidence interval
E+P HT	Estrogen and progestin hormone therapy
EMA	European Medicines Agency
EU	European Union
FDA	Food and Drug Administration
HIRD	HealthCore Integrated Research Database
HR	Hazard ratio
IEC	Independent ethics committee
IR	Incidence rate
IRB	Institutional review board
IRR	Incidence rate ratio
IUD	Intrauterine device
MDCD	MarketScan Medicaid database
MDCR	Medicare
MI	Myocardial infarction
N	Number

Abbreviation	Definition
N/A	Not applicable
NDC	National drug code
NMSC	Non-melanoma skin cancer
ORD	Optum Research Database
PASS	Post-authorization safety study
PHI	Protected health information
PS	Propensity score
PPV	Positive predictive value
RR	Relative risk
SAP	Statistical analysis plan
SERM	Selective estrogen receptor modulator
TIA	Transient ischemic attack
US	United States
VTE	Venous thromboembolism

3. INVESTIGATORS

Principal Investigator(s) of the Protocol

Name, degree(s)	Title	Affiliation
Renu Garg, PhD, MPH	Safety Surveillance Research Scientist	Pfizer Inc.
Sarah Hoffman, PhD, MS, MPH	Senior Researcher, Safety & Epidemiology	HealthCore, Inc.
Daniel Beachler, PhD, MHS	Director, Safety & Epidemiology	HealthCore, Inc.
Florence Wang, ScD	Executive Director, Epidemiology	Optum
Cheryl McMahill-Walraven, MSW, PhD	Director, Informatics	Healthagen (currently known as CVS Health Clinical Trial Services as of 01 November 2020)

4. OTHER RESPONSIBLE PARTIES

Responsible Party Name and Affiliation	Role in the study
Nicole Fournakis, MPH HealthCore	Project Manager
Kimberly Daniels, PhD, MS HealthCore	Analyst
Samuel Governor, MD, MPH HealthCore	Analyst
Carla Brannan, BS Healthagen (currently known as CVS Health Clinical Trial Services as of 01 November 2020)	Project Manager
Patricia Shuminski, AS Healthagen (currently known as CVS Health Clinical Trial Services as of 01 November 2020)	Analyst
Dingwei Dai, MD, PhD Healthagen (currently known as CVS Health Clinical Trial Services as of 01 November 2020)	Analyst
Alison Edwards, MStat Healthagen (currently known as CVS Health Clinical Trial Services as of 01 November 2020)	Statistician
Ryan Seals, MS, ScD Optum	Epidemiologist
Najat Ziyadeh, MPH Optum	Senior Scientist
Nicole Brooks, MSPM, PMP Optum	Epidemiology Project Manager

5. MILESTONES

Milestone	Planned date	Actual date	Comments
Date of institutional review board (IRB) approval of Protocol	Not Applicable	20 February 2015	
Start of data collection	01 November 2015	01 November 2015	This is a cohort study and the source data have been accruing since *Duavee* became available in the US in May 2014. In accordance with EU Pharmacovigilance Guidance (Module VIII), the start of a PASS using secondary data is defined to be when the data for the analytic dataset are first extracted.
Interim Report 1	31 March 2016	31 March 2016	
Interim Report 2	31 March 2017	31 March 2017	
Interim Report 3	31 March 2018	31 March 2018	
Interim Report 4	31 March 2019	31 March 2019	
End of data collection	01 November 2019	31 August 2020	
Registration in the EU PAS register	01 October 2015	01 October 2015	
Final report of study results	31 March 2021	26 March 2021	

6. RATIONALE AND BACKGROUND

Menopause is the cessation of ovulation resulting in reduced ovarian production of estradiol (estrogen) and cessation of menstruation.[1] Clinically, menopause is defined as the absence of a menstrual period for at least 12 months.[2] The time leading up to menopause is referred to as perimenopause or menopausal transition and is characterized by fluctuating estrogen levels, shortened, prolonged, or absent menses, hot flashes, night sweats, insomnia, vaginal dryness, painful intercourse, frequent urination, and urinary tract infections.[3] The median age at menopause among White women in industrialized nations is approximately 51 years, with perimenopause onset at 47.5 years.[1]

Vasomotor symptoms (hot flashes and night sweats) are the most common symptoms of menopausal transition, afflicting as many as 80% of women who undergo the transition, and last a median of 7.4 years.[4] To alleviate these symptoms, many women resort to exogenous estrogen formulations, collectively referred to as hormone therapy.[5,6]

Hormone therapy formulations for vasomotor symptoms associated with natural menopause in women with an intact uterus often include progestin (synthetic progesterone) to protect patients from the increased risk of endometrial cancer associated with taking unopposed estrogen.[7] However, progestin may be avoided by women concerned about breast cancer risk[8-10] or who experience side effects from progestin, such as abnormal uterine bleeding.[11] Due to these concerns, there was a market need for an alternative approach to estrogen opposition in hormone therapy.

On 03 October 2013, the oral conjugated estrogens/bazedoxifene (CE/BZA), Duavee®, was authorized in the United States (US) for the treatment of moderate to severe vasomotor symptoms associated with menopause and prevention of postmenopausal osteoporosis[12] in women with a uterus.[13] In the European Union (EU), the indication for CE/BZA is slightly different; it is not indicated for the prevention of osteoporosis, and the language stipulates that CE/BZA is for women in whom "treatment with progestin-containing therapy is not appropriate."[14] Instead of progestin, CE/BZA relies on bazedoxifene to oppose estrogen. Bazedoxifene is a selective estrogen receptor modulator (SERM) developed for osteoporosis prevention and treatment.[15]

To date, a variety of combination hormone therapy products for menopause are available on the US market (Table 1).[16] Most products are available in oral form, contain estradiol, norethindrone acetate, and/or were approved in the late 1990s or early 2000s. Two transdermal patch products are available.

Table 1. US Food and Drug Administration (FDA) approved combination hormone therapy products available by prescription in the United States, by estrogen and progestin type

Commercial Name	Estrogen Type	Progestin Type	Route	Initial US Approval*
Duavee	Conjugated estrogen	N/A†	Pill	2013
Prempro & Premphase	Conjugated estrogen	Medroxyprogesterone	Pill	1995
Activella	Estradiol	Norethindrone acetate	Pill	1998
Combipatch	Estradiol	Norethindrone acetate	Patch	1998
Femhrt	Ethinyl estradiol	Norethindrone acetate	Pill	1999
Angeliq	Estradiol	Drospirenone	Pill	2005
Climara Pro	Estradiol	Levonorgestrel	Patch	2003
Prefest & Ortho-Prefest	Estradiol	Norgestimate	Pill	1999

Abbreviations: N/A, not applicable; US, United States.
*Source: https://www.accessdata.fda.gov/scripts/cder/daf/
†Bazedoxifene instead of progestin

This study collected real-world data on the risks of various endpoints among populations for which Duavee (i.e., CE/BZA) is prescribed (named Duavive in the EU). Because Duavee is an estrogen-containing product, endometrial hyperplasia and endometrial cancer are relevant safety events and were designated as primary safety endpoints, while other safety events were selected as secondary endpoints.

During the first four years after introduction of CE/BZA in the US (01 May 2014 through 31 May 2018), the rate of uptake of new users of this medication in the single database originally intended for this study was lower than projected. To conduct a meaningful assessment of the risk of the primary safety events, the study was expanded to include four additional US databases to minimize the time needed to achieve an adequate study size.

This non-interventional study was designated as a Post-Authorization Safety Study (PASS) and was a commitment to the European Medicines Agency (EMA).

7. RESEARCH QUESTION AND OBJECTIVES

The overall aim of this PASS was to monitor the safety profile of Duavee (CE/BZA) in comparison to estrogen and progestin combination hormone therapy (E+P HT). Towards this end, the following primary and secondary objectives were completed:

Primary Objective:

To estimate the incidence and compare the risks of endometrial hyperplasia and endometrial cancer among postmenopausal women initiating CE/BZA and postmenopausal women initiating E+P HT during the first five years of CE/BZA availability in the US (2014-2019).

Secondary Objectives:

To estimate the incidence and compare the risks of the following safety outcomes among postmenopausal women initiating CE/BZA and postmenopausal women initiating E+P HT:

- Venous thromboembolism (VTE)
- Myocardial infarction (MI)
- Stroke (including transient ischemic attack [TIA])
- Breast cancer
- Ovarian cancer
- Thyroid cancer
- Renal cancer and adenoma
- Gastrointestinal cancer
- Any cancer (any malignant neoplasm, including cancers listed above and excluding non-melanoma skin cancer [NMSC]).

8. AMENDMENTS AND UPDATES

No amendments to the Protocol have been made since the EMA's final Assessment Report (AR), dated 21 April 2020. All Protocol amendments prior to the final AR are described in the table below. In response to the EMA's request to provide information on the proportion of patients that were lost-to-follow-up after turning 65 for the individual databases, a description of this analysis and the data in the pooled analysis (Section 10.5.2 and **Appendix 8f, Table 14**) have been included in the final report.

Table 2. Amendments to the Protocol

Amendment number	Date	Substantial or administrative amendment	Protocol section(s) changed	Summary of amendment	Reason
1	28 November 2017	Administrative	**7.3.2**	An additional end point of any cancer, excluding basal cell carcinoma, was added, in addition to the original end point of any cancer. Subsequently, this end point was further revised to "any cancer, excluding non-melanoma skin cancer (NMSC)".	Non-melanoma skin cancer (which includes basal cell carcinoma) is common and non-fatal with a behavior similar to benign tumor, it is often excluded when assessing all-cause cancer in epidemiologic studies as it is excluded from US cancer registries including the US Surveillance, Epidemiology, and End Results (SEER) national registry.
2	15 February, 2018	Administrative	**7.3.4**	Covariate updated from "cerebrovascular disease" to "cardiovascular or cerebrovascular disease".	The codes in the appendix indicated both cardiovascular and cerebrovascular disease, which are confounders to secondary outcome – stroke.
3	15 February 2018	Administrative	**7.3.4**	"Antimycobacterials" was removed from and "dermatologicals" was added in the list of covariates.	Variable label update was made to reflect consistency with the analyses.
4	31 March 2018	Administrative	**7.3.3;** **10.1**	Mortality Supplement to Final Report was removed from Milestone table and text references were removed.	Previously, a Mortality Supplement to the Final Report was planned for submission by March 31st, 2021. Because the NDI data lag has shortened substantially since the protocol was written, this supplemental report is no longer necessary.
5	30 April 2019	Substantial	Throughout entire Protocol amendment (highlighted in red font)	The inclusion of two additional Research Partners (Healthagen and Optum) has been documented, the study power calculations have been updated, and the methodology and analysis plan have been augmented to describe the aggregation of results	Current trends indicate this PASS was not able to accumulate enough patients to conduct a meaningful assessment of the risk of the primary safety events (endometrial hyperplasia and endometrial cancer) among new users of the study drugs in the near

Table 2. Amendments to the Protocol

Amendment number	Date	Substantial or administrative amendment	Protocol section(s) changed	Summary of amendment	Reason
				across study partners. All codes have been removed from the Protocol and were included in a statistical analysis plan (SAP), which was co-developed by the three research partners and Pfizer Inc. Sections on National Death Index (NDI) linkage, mortality analyses, and algorithm validation have been removed. Description of inclusion/exclusion criteria and analyses has been updated to improve clarity and reproducibility between research partners. The study period has been extended to 2014-2019, and the expected date of the final study report has extended. The previous final study report has been changed to be the 4th interim report.	future using only the HIRD. Therefore, the study was expanded to include four additional databases to accelerate the time needed to achieve an adequate study size. Given that some of these databases do not have the ability to link to the NDI for accurate mortality assessment, the all-cause mortality measures in this study were excluded.
6	23 August 2019	Substantial	Throughout the entire Protocol amendment	The inclusion of two additional databases, analyzed by the HealthCore team (thus leading to five databases in this study). Start and end of data collection dates were amended for consistency with definitions in EMA's Guideline on Good Pharmacovigilance Practices (GVP) Module VIII – Post-authorisation safety studies. Further information was provided regarding the proposed meta-analysis.	The MarketScan commercial database was included to increase total sample size. The MarketScan Medicaid database was included based on the recommendation of the Pharmacovigilance Risk Assessment Committee (PRAC) to include patients without commercial health insurance and those who are covered under public programs. The inclusion of these additional databases modestly increased the study size and improved the power

Table 2. Amendments to the Protocol

Amendment number	Date	Substantial or administrative amendment	Protocol section(s) changed	Summary of amendment	Reason
				Added updated information on the risk estimates that can be ruled out. Addressed whether women may be included in several databases. Minor edits were made throughout for clarity.	to detect differences in endometrial cancer and hyperplasia risk between CE/BZA new users and E+P HT new users. Additional changes were made to address comments in PRAC Response to Third Interim Report.

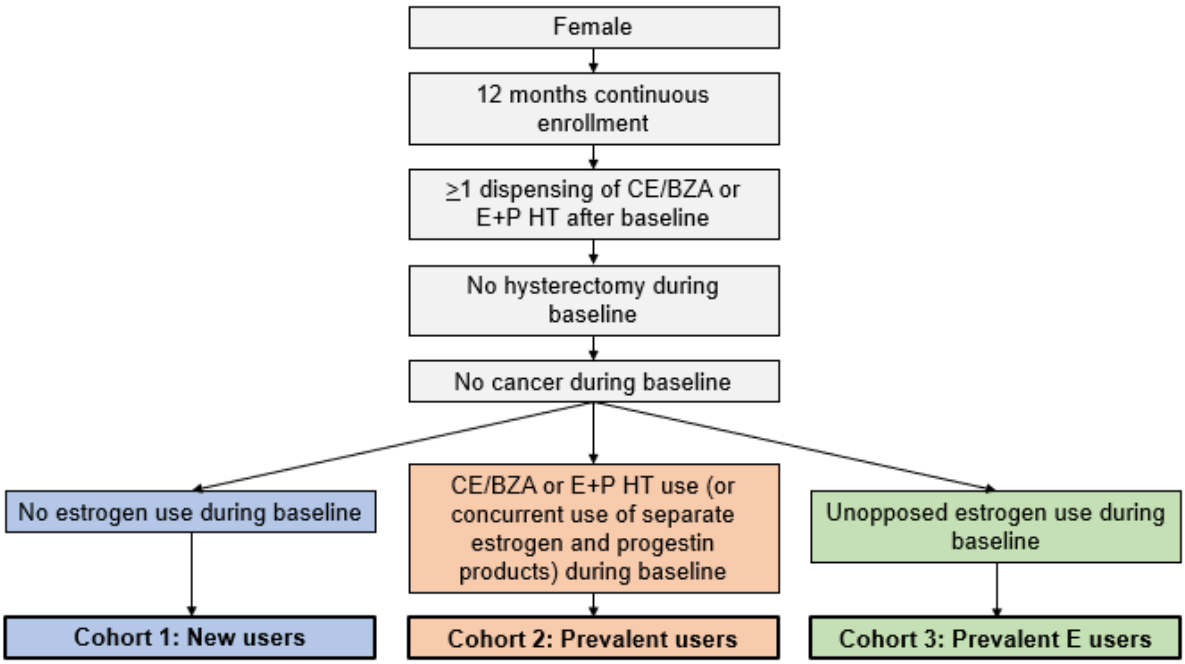
9. RESEARCH METHODS

9.1. Study design

A multi-database cohort study in five US healthcare claims databases was conducted to examine the risks of endometrial cancer and endometrial hyperplasia in CE/BZA users versus E+P HT users. The primary analysis consisted of a new-user, active comparator design, while secondary analyses examined prevalent users. CE/BZA was contrasted with an active comparator group consisting of oral, topical, and transdermal E+P HT commonly used for hormone therapy (Section 9.4.1). Primary outcomes included endometrial cancer and endometrial hyperplasia. Secondary outcomes included nine safety events: three acute cardiovascular outcomes and six cancer outcomes. Incidence rates and hazard ratios were reported for all study outcomes.

Study inclusion/exclusion criteria and the baseline period are described in Figure 1. Three cohorts were constructed based on estrogen exposure during the baseline period (Figure 2). Hazard ratios were not estimated for the prevalent user cohorts due to concern regarding prevalent user bias,[17-19] however, propensity score (PS)-matched incidence rates are presented.

Figure 1. Inclusion/exclusion criteria and creation of three cohorts*

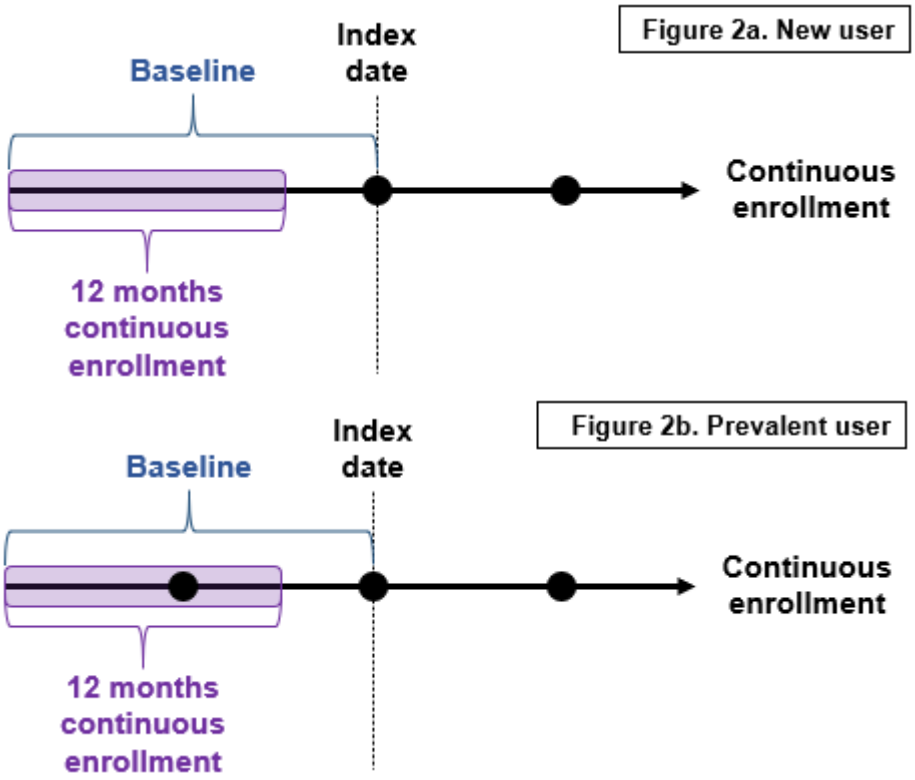


Definitions/Abbreviations: Baseline, defined as the period of continuous enrollment prior to index date (minimum of 12 months); CE/BZA, conjugated estrogens/bazedoxifene; E, unopposed estrogen; E+P HT, estrogen and progestin combination hormone therapy.

*Order of steps varied slightly between databases. Inclusion/exclusion criteria specific to study outcome were also applied: (1) no baseline endometrial hyperplasia diagnosis for the endometrial hyperplasia analyses, and (2) no cardiovascular outcomes in the last six months of the baseline period for the cardiovascular outcomes analyses. For all

analyses, women with both a CE/BZA and E+P HT dispensing on the index date were excluded, and women with CE/BZA or E+P HT and unopposed estrogen dispensing on the index date were also excluded.

Figure 2. Schematic depicting index date determination*



*Black dots represent dispensings. The first fill after 12 months of continuous enrollment during the study period, i.e., 01 May 2014 to 31 August 2019 (or 31 December 2018 for MarketScan Medicaid or 31 July 2019 for MarketScan CCAE/Medicare) serves as the *index date*. In Figure 2a, the patient would be included in cohort 1 (new user cohort), and in Figure 2b, the patient would be included in either cohort 2 or 3 (prevalent user cohorts) due to their dispensing during the 12 months of continuous enrollment, which occurs before the index date by definition. Depending on how far into the study period the index date occurs, the 12 months continuous enrollment could have occurred before and/or during the study period. The earliest possible index date was 01 May 2014.

9.2. Setting

Databases. The five study databases included (1) Healthagen (currently known as CVS Health Clinical Trial Services as of 01 November 2020), which used Aetna's Sentinel Common Data Model, (2) HealthCore Integrated Research Database (HIRD), (3) MarketScan Commercial Claims & Encounters (CCAE) & Medicare Supplemental database (4) MarketScan Medicaid database (MDCD), and (5) Optum Research Database (ORD). These databases are further described in this study's Protocol (**Appendix 2, Section 7.5**).

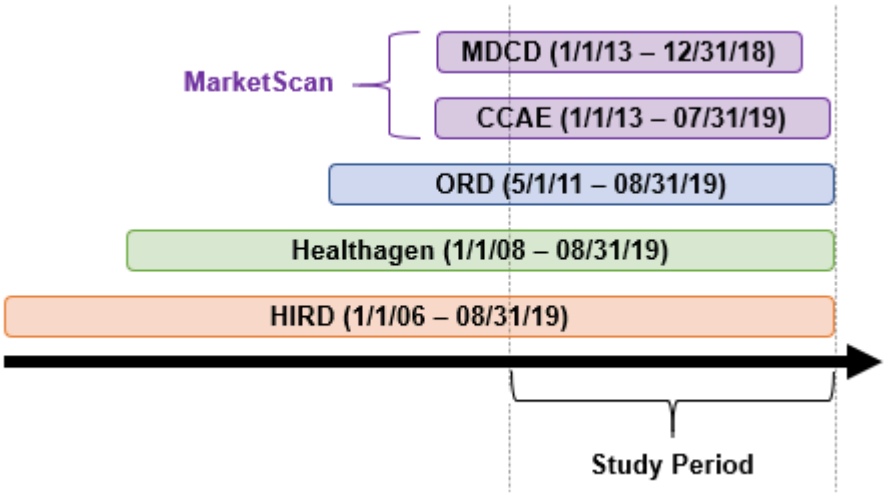
While it is possible for patients to be double-counted across databases if they switched insurance plans during the study period, their events and exposure-classified person-time could not be double-counted. Person-time serves as the basis of the main analyses in this

study. However, it is possible that the MarketScan CCAE & Medicare Supplemental database includes some duplicate/contemporaneous claims and person-time from three of the other databases (i.e., the non-Medicaid databases). For this reason, a sensitivity analysis was carried out excluding the MarketScan CCAE & Medicare Supplemental database (Section 10.5.1.3).

Study period. The study period, i.e., the time period during which an *index date* (Figure 2) could occur, was 01 May 2014 to 31 August 2019 (or 31 December 2018 for MarketScan Medicaid or 31 July 2019 for MarketScan CCAE/Medicare;Figure 3). Patients with an index date on the last day of the study period were excluded due to inadequate follow-up time.

Covariate assessment period. Figure 3 describes dates of data availability for covariate ascertainment. For certain relevant variables (Statistical Analysis Plan **[SAP] Appendix 4, Table 7**), an all-available lookback approach was used, meaning that all available pre-index date data were used for each patient, with an inherent 12-month minimum due to the continuous enrollment requirement (Figure 1). Other variables were assessed with a fixed, six or 12-month lookback period (**Appendix 4, Table 7**).

Figure 3. Dates of data availability for covariate assessment, by study database



Definitions/Abbreviations: CCAE, MarketScan Claims & Commercial Encounters & Medicare Supplemental Database; HIRD, HealthCore Integrated Research Database; MDCD, MarketScan Medicaid Database; ORD, Optum Research Database.

Follow-up period. Follow-up for each patient began on the day after the index date and lasted until death, disenrollment, the end of the study period (Figure 3), hysterectomy, treatment switch (defined as E+P HT to CE/BZA or CE/BZA to E+P HT; switching within E+P HT drug list was not censored), occurrence of the specific outcome under investigation, or censoring specific to that outcome (see next paragraph). For acute outcomes (i.e., endometrial hyperplasia, VTE, MI, and stroke/TIA), follow-up was further limited to "current use" treatment episodes.

Outcome-specific censoring. Outcome-specific censoring is described in the SAP **(Appendix 4, Table 3)**. In brief, cancer outcomes were not censored by any additional conditions while endometrial hyperplasia, VTE, MI, and stroke/TIA were censored by any cancer diagnosis or end of current treatment episode.

Statistical Software. Data management and analyses took place in 2019-2020 using Statistical Analysis System (SAS) Enterprise Guide version 7.15 (SAS Institute, Cary, NC, US) and SAS 9.4 (SAS Institute Inc., Cary, NC). Analyses were completed by HealthCore (HIRD and MarketScan databases), Pfizer (MarketScan databases), Optum (ORD), and Healthagen. Pooled analyses were conducted by HealthCore using the *metafor* and *meta* packages in R.[20,21]

9.3. Subjects

Inclusion/exclusion criteria are depicted in Figure 1. In brief, the study population included female participants of any age with a dispensing of CE/BZA or E+P HT after at least 12 months of continuous enrollment without pre-index cancer or hysterectomy. For comparative analyses (Cohort 1), the analytic population consisted exclusively of PS-matched patients. Unmatched patients were not included in the comparative analyses. Section 9.9.2 describes the PS matching approach used.

9.4. Variables

Variables were classified as either exposures (Section 9.4.1), outcomes (Section 9.4.2), or covariates (potential confounders; Section 9.4.3) per standard epidemiologic terminology.[22]

9.4.1. Exposure assessment

Patients were considered exposed to CE/BZA if they had at least one pharmacy claim with a National Drug Code (NDC) for CE/BZA (**Appendix 4, Table 1**). Patients were considered exposed to E+P HT if they had at least one pharmacy claim with an NDC for oral, topical, or transdermal E+P HT (**Appendix 4, Table 1**). E+P HT drugs included:

- Conjugated estrogens/medroxyprogesterone acetate
- Estradiol and norethindrone acetate
- Norethindrone acetate-ethinyl estradiol
- Estradiol/drospirenone
- Estradiol/levonorgestrel
- Estradiol/norgestimate
- Estradiol-estriol-progesterone micronized cream.

Treatment group assignment. Exposure classification was determined at the first fill after 12 months of continuous enrollment. If a patient had a prior fill of CE/BZA, E+P HT, or unopposed estrogen prior to or during the required 12-month continuous enrollment period, they were placed into one of the two prevalent user cohorts (Figure 1 and Figure 2). Patients

with no prior fill of CE/BZA, E+P HT, or unopposed estrogen were included into the new user cohort (Figure 1 and Figure 2).

Treatment episodes. CE/BZA and E+P HT treatment episodes were constructed by concatenating consecutive dispensing days supplies, i.e., fill date + days supply, allowing for treatment gaps of 30 days (Optum allowed up to 32 days). If a patient had a dispensing of the same study drug that occurred before the end of a treatment episode, the treatment episode was extended by the number of overlapping days (**Appendix 4, Section 3.1.1**). For E+P HT users, switching between E+P HT products constituted a continuation of E+P HT use.

9.4.2. Outcome assessment

Primary Outcomes

Primary outcomes included endometrial cancer and endometrial hyperplasia. Endometrial hyperplasia and endometrial cancer cases were identified through previously published, validated algorithms, [23,24] developed in E+P HT users in the HIRD. Endometrial cancer was identified by the presence of ≥1 inpatient hospitalization with a principal diagnosis of endometrial cancer or ≥2 outpatient or emergency room visits on different dates with an endometrial diagnosis in any position. This endometrial cancer algorithm had a demonstrated positive predictive value (PPV) of 90.8% (95% confidence internal [CI] 86.9 – 93.6).[24] Endometrial hyperplasia was defined using a predictive model algorithm with >25 predictors (**Appendix 4**) with a demonstrated PPV of 80% (95% CI 77%-88%) in the HIRD validation study.[18] These algorithms are further described in Section 3.2.1 of the SAP (**Appendix 4**).

Secondary Outcomes

Secondary outcomes included nine safety events – three acute cardiovascular outcomes and six cancer outcomes. Venous thromboembolism (VTE) cases were identified by the presence of ≥1 inpatient hospitalization with a principal diagnosis of VTE (**Appendix 4**). Myocardial infarction (MI) and stroke/TIA cases were identified by the presence of either (a) ≥1 inpatient hospitalization with a principal diagnosis code for the condition with a length of stay of ≥3 days, or (b) ≥1 inpatient hospitalization with a principal diagnosis of the condition and a discharge status of death, or (c) ≥1 emergency room visit with a principal diagnosis code for the condition (**Appendix 4**). Cancer outcomes required either (a) ≥1 inpatient hospitalization with a principal diagnosis of the cancer under study, or (b) ≥2 outpatient or emergency room visits on different dates with the relevant diagnosis in any position (**Appendix 4**). The breast cancer definition also included inpatient hospitalizations with a diagnosis of breast cancer in any position (**Appendix 4**).

9.4.3. Covariate assessment

Covariates and their time frames are reported in the SAP **(Appendix 4, Table 7)**. In brief, covariates included age, sex, census region, total time in health plan prior to and including index date, calendar year of index date, pre-specified comorbidities and medications, and the 25 most common comorbidities and medications in the pre-matched sample by database and

cohort (**Appendix 4, Table 7**). Endometrial hyperplasia as a covariate was defined as the presence of an endometrial hyperplasia diagnosis code on any claim in any position. Covariates were entered into PS models and patients were matched on PS (Section 9.9.2).

9.5. Data sources and measurement

The study included five US insurance claims databases (Table 3). Each database captures the healthcare experience of persons from across the US who are covered by commercial or public health insurance. These data are composed of individuals' health insurance claims and enrollment information. Patient eligibility criteria, covariates for statistical control, and statistical methods for analysis were applied identically across the databases except where noted (Section 9.9.2.1). Differences in data availability, variable ascertainment, and execution of analyses between research partners and databases are noted in Section 9.9.2.1.

Table 3. Characteristics of the five study databases

Database	Payor or Servicer	Includes ASO Patients?	Includes Medicare Patients?	Data Availability for this project
HIRD	Private (Anthem)	Yes	Advantage	01/2006 – 08/2019
MarketScan CCAE-MDCR	Private (Mixed, ES)	NA	Supplemental	01/2013 – 07/2019
MarketScan Medicaid	Public (Medicaid)	NA	Yes	01/2013 – 12/2018
Healthagen*	Private (Aetna)	No	Advantage	01/2008 – 08/2019
ORD	Private	Yes	No	05/2011 – 08/2019

Abbreviations: ASO, administrative services only; CCAE-MDCR, commercial claims and encounters-Medicare; ES, Employer-sponsored; HIRD, HealthCore integrated research database; ORD, Optum research database.
*Utilized Aetna's Sentinel Common Data Model.

9.6. Bias

CE/BZA and E+P HT users may differ in ways that are related to the outcomes under study, leading to confounded comparisons between the two treatments. To address confounding related to demographic characteristics, pre-existing comorbidities, and prior or current use of other prescription drugs, patients were matched to similar patients using a PS-based approach (Section 9.9.2).

To address the potential for misclassification of endometrial cancer and endometrial hyperplasia outcomes by exposure status, a sensitivity analysis was carried out to examine the findings for different sensitivity and PPV scenarios (Section 9.9.4 & Section 10.5.1.3).

9.7. Study size

This study relied on secondary data and included all eligible exposed patients (i.e., there was no sampling). The final, analytic population included a total of 75,455 patients (18,417 CE/BZA and 57,038 E+P HT users) from five databases (Section 10.1;Table 5), including 44,414 new users (10,596 CE/BZA).

In the Protocol (**Appendix 2**), the assumptions were a 4:1 matching ratio and background rates of 81.4 cases per 100,000 person-years for endometrial cancer and 142.9 cases per 100,000 person-years for endometrial hyperplasia, an alpha level of 0.05, and an average duration of follow-up of one year per patient. A study size of approximately 13,698 new users of CE/BZA was estimated to have 80% power to detect a hazard ratio of 1.8 or higher for endometrial hyperplasia and 2.1 or higher for endometrial cancer. The number of new CE/BZA users in this study was lower than in the Protocol's projections due to the application of additional exclusion criteria (prior use of any estrogen).

9.8. Data transformation

Detailed methodology for data transformations, particularly complex transformations (e.g., many raw variables used to derive an analytic variable), are documented in the SAP, which is dated, filed and maintained by the sponsor (SAP; **Appendix 4**).

9.9. Statistical methods

9.9.1. Main summary measures

- Incidence rates (IRs) were calculated as the total number of patients with the outcome of interest divided by the total person-time at risk for that outcome (see outcome-specific censoring, Section 9.2). Covariates were accounted for by PS matching (Section 9.9.2.1.2).

- Incidence rate ratios (IRR) were calculated as the IR among CE/BZA users divided by the IR among E+P HT users [referent] for each outcome. Covariates were accounted for by PS matching (Section 9.9.2.1.2).

- Hazard ratios (HR) were calculated by constructing Cox proportional hazards models for each cohort to estimate the HR for each outcome. Kaplan-Meier plots were used to graphically assess the proportional hazards assumption in the two groups. Covariates were accounted for by PS matching (Section 9.9.2.1.2).

9.9.2. Main statistical methods

To identify E+P HT users who were comparable to patients in the CE/BZA cohort, the probability of initiating CE/BZA versus E+P HT was estimated for each patient. HealthCore and Research Partners (Pfizer, Optum, and Healthagen) identified *a priori* all covariates to be pre-specified in the PS models with the exception of the 25 most common diagnoses and procedures which could vary by database, contingent on the data. In each of the five databases, three separate PS-matched cohorts (Figure 1) were developed for this study. Logistic regression was used to estimate PS as the probability of receiving a CE/BZA dispensing given the specified covariates (Section 9.4.3**; Appendix 4, Table 7**). Frequency matching without replacement was implemented unique to each database with specific parameters reported in Section 9.9.3 and Table 4. Absolute standardized differences were computed to assess covariate balance, with an *a priori* threshold of 0.1.[25,26]

Cox models are described in Section 9.9.2.1.3.

Pooling of study results is described in Section 9.9.2.2.

9.9.2.1. Database-specific methods

9.9.2.1.1. Propensity score (PS) modeling

Cohort specific models: For each database, three separate sets of PS, one for each cohort, were estimated for this study. For ORD, this was implemented via the estimation of one model that included patients in all three cohorts, with indicator variables identifying membership in the three cohorts.

Excluded variables due to low counts: In the **HIRD** analyses, the following baseline variables were excluded from the final PS model due to zero cell counts: baseline copper intrauterine device (IUD) and progestin IUD utilization, vaginal progestin, combined oral contraceptives, progestin-only oral contraceptives, topical progestin, history of breast/ovarian/endometrial malignancy, thrombophilias, testosterone, SERMs, number of estrogen prescriptions. In **MarketScan CCAE & Medicare Supplemental** analyses, the following baseline variables were excluded from the final PS model due to zero cell counts: vaginal progestin, combined oral contraceptives, number of estrogen prescriptions, and copper IUD (Cohort 2 [prevalent users] only). In the **MarketScan Medicaid** analyses, the following baseline variables were excluded from the final PS model due to zero cell counts: baseline copper IUD and progestin IUD utilization, number of estrogen prescriptions, topical progestin, vaginal progestin, combined oral contraceptives, and due to model non-convergence, corticosteroids, macrolides, azithromycin, glucocorticosteroids. In the **ORD** analyses, baseline copper IUD utilization was excluded from the final PS models due to zero cell counts.

Excluded variables due to pairwise correlations: Optum assessed pair-wise correlations for all covariate pairs, and correlations above 0.8 were flagged for the removal of one of the variables. After review, the following variables were excluded from Optum PS modeling due to high correlations with other variables in the model: topical hormone treatments, sedatives/hypnotics, progestin IUD, macrolides, antifungals, oral contraceptives, lipid lowering agents, antivirals, oral corticosteroids, and essential hypertension. These high correlations were generally due to the empirically-identified 25 most common medications and diagnoses being highly similar in definition to pre-specified covariates. Thus, the exclusion of these covariates does not imply that the PS model did not capture data relevant to the covariate.

Additional variables: In addition to the a priori variables, Optum included the following variables in their PS model, all (other than the time from start of study period variable) of which were assessed in the 183 days prior to and including the index date: Any emergency room visit (yes/no), time from start of study period (01 May 2014) to index date, number of 3-digit diagnoses codes, number of inpatient stays, number of procedures, number of unique procedures, number of drugs dispensed, number of physician visits, total healthcare costs.

Interaction terms: After the assessment of imbalance between study exposure groups (CE/BZA and E+P HT), Optum included interaction terms between cohort membership (cohort 1, 2 or 3) and the following baseline covariates in the final PS model: family history of cancer, breast pain or lump, and renal disease.

9.9.2.1.2. Propensity Score (PS) matching

The following table summarizes the PS matching methods used in each database.

Table 4. Propensity score (PS) matching methods by study database

Database	Matching Interval	Matching Ratio	Notes
HIRD			
Cohort 1	Decile	3:1	N/A
Cohort 2	Decile	3:1	N/A
Cohort 3	N/A	N/A	Matching not feasible due to low patient count.
MarketScan CCAE-MDCR			
Cohort 1	Quintile	3:1	
Cohort 2	Decile	3:1	
Cohort 3	Quartile	1:1	
MarketScan Medicaid			
Cohort 1	Quintile	5:1	
Cohort 2	Quartile	1:1	
Cohort 3	N/A	N/A	Matching not feasible due to low patient count.
Healthagen*			
Cohort 1	Nearest neighbor matching within PS calipers of 0.2 standard deviation of the logit of the PS.	4:1	Copper and hormonal IUDs combined into one variable.
Cohort 2	Nearest neighbor matching within PS calipers of 0.2 standard deviation of the logit of the PS.	4:1	ASD > 0.1 for one or more variables. Copper and hormonal IUDs combined into one variable.
Cohort 3	N/A	N/A	Matching not feasible due to low patient count.
ORD			
Cohort 1	Decile	4:1	Included Optum standard variables†
Cohort 2	Quintile	3:1	Included Optum standard variables†
Cohort 3	Quintile	2:1	Included Optum standard variables†

Abbreviations: ASD, absolute standardized difference; CCAE-MDCR, commercial claims and encounters-Medicare; HIRD, HealthCore integrated research database; IUD, intrauterine device; N/A, not applicable; ORD, Optum research database; PS, propensity score.
Cohort 1: New Users; Cohort 2: Prevalent Users; Cohort 3: Prior Unopposed Estrogen
*Used Aetna's Sentinel Common Data Model.
†Any emergency room visit (yes/no), Time from start of study period (01 May 2014) to index date, Number of 3-digit diagnoses codes, Number of inpatient stays, Number of procedures, Number of unique procedures, Number of drugs dispensed, Number of physician visits, Total healthcare costs.

9.9.2.1.3. Cox proportional hazards modeling

Stratification: For all Cox proportional hazards models, Healthagen stratified on the PS-matched pairs.

Addition of variables to address residual imbalance after PS matching: For the ORD analyses, indicators for year of cohort entry were included in all final comparative models due to moderate imbalances (i.e., absolute standardized difference ≥0.1) that persisted after PS matching.

9.9.2.2. Pooled analyses and heterogeneity testing

DerSimonian and Laird's Random Effects model was used to pool the effect sizes for IR, IRR, and HR. The Cochran's Q test and i^2 values were used to examine and quantify the heterogeneity of the results between databases. The i^2 value represents the proportion of the observed variance not attributable to sampling (random) error.[27] Since the meta-analysis involves a small number of studies, and Cochran's Q test is insensitive and has relatively low power, the heterogeneity p-value was set at < 0.1 to indicate the presence of heterogeneity.[28] The i^2 value and the Q test value, using the formula below, can be used to estimate the precision interval for the possible range of values that the true effects for the primary outcomes are likely to have 95% of the time.[27]

$$i^2 = \frac{\textbf{variance in true effect}}{\textbf{variance in observed effect } (\boldsymbol{Q})}$$

A continuity correction of 0.5 was used to adjust zero cell frequencies in the meta-analysis for IR and IRR. All meta-analyses were conducted in R using the *metafor* and *meta* packages.[20,21]

9.9.3. Missing values

In claims data there are no indicator variables (yes/no) for specific diseases or treatments. Diseases and treatments are ascertained by presence of diagnosis, procedure, or medication codes on claims. The absence of a diagnosis, procedure, or medication code does not imply absence of the condition or treatment; it simply implies that the condition or treatment was not relevant to the billing for that healthcare encounter. Therefore, there is no way to identify "missing" values for presence or absence of a condition or treatment in claims. To allow for analysis, this study assumed that absence of a code or chart note regarding a diagnosis, procedure, or medication implies its absence or irrelevance to a given patient at a given time. While missing claims cannot be quantified, it is possible to quantify missing values for components of claims or enrollment data. Missing demographic variables were quantified and reported in results tables either as a row for "unknown" or "missing" (for categorical or discrete variables) or as a footnote (for variables reported as continuous). Region was not available in the MarketScan Medicaid database and this database was excluded from combined counts for region.

9.9.4. Sensitivity analyses

To explore the potential for effect modification by age, duration of use, timing of use, route of administration, or prior osteoporosis, this study stratified the endometrial cancer and endometrial hyperplasia analyses in new users (Cohort 1) by each of these covariates (**Appendices 8a-e**, **Tables 8b & 8c**).

To address how censoring criteria may have affected study findings, additional sensitivity analyses were conducted varying the censoring criteria. Additional analyses were performed in new users (Cohort 1) removing censoring due to a change in treatment during follow-up, removing censoring due to a cancer diagnosis during follow-up for non-cancer outcomes, restricting to first new use treatment episode, including non-melanoma skin cancer in the all cancer outcome, and evaluating acute outcomes with an expanded timeframe (**Appendices 8a-e, Table 9**).

MarketScan CCAE & Medicare Supplemental database may include patients who are also represented in the HIRD, ORD, or Healthagen data. To address the potential overlap in events and person-time, pooled estimates were re-calculated excluding MarketScan CCAE & Medicare Supplemental data (**Appendix 8f, Table 12**).

To address the potential for misclassification of endometrial cancer or endometrial hyperplasia by exposure status, quantitative bias analyses were carried out to examine the findings across different sensitivity and PPV scenarios (**Appendix 8f, Table 13**). The methods used for this analysis are further described, along with the results, in Section 10.5.1.3

9.9.5. Amendments to the statistical analysis plan

The SAP was developed by the project team prior to implementation of the data management and analyses phase of this project (**Appendix 4**). While amendments were not made to the written SAP, several minor modifications were made to the original Table Shells (**Appendix B of SAP [Appendix 4]**) and Code Lists Appendices (**Appendix A & D of SAP [Appendix 4]**) during the data management and analysis phases of the project. These minor amendments were incorporated to clarify the language of the table shell footnotes, combine code lists for select variables, and add codes for other variables. The final code list document utilized for all five databases is provided in Appendix A & D of the SAP (**Appendix 4**), and the final results for each database are provided in **Appendices 8a-e** of this report.

Per a request by the EMA, a sensitivity analysis was added to examine the amount of censoring in each database due to US Medicare enrollment (i.e., when turning age 65).

For the quantitative bias analysis, the SAP had described using various sensitivity and specificity estimates to assess potential misclassification of endometrial cancer and endometrial hyperplasia outcomes. However, the formulae and table shells have been updated to use sensitivity and PPV, as PPV rather than specificity was estimated by the HIRD validation studies for these two outcomes. This approach more closely aligns with the approved Protocol which specified that the bias analyses would be guided by the HIRD

validation study findings and cited the Brenner (1993) paper for use of sensitivity and PPV in the formula (**Protocol Section 7.8.4**).[29].

9.10. Quality control

All results tables were carefully checked for plausibility and internal consistency by study principal investigators DB & SH. Additionally, findings from each database were examined in the context of findings from the other four databases. When differences in findings or distribution of study variables could not be easily explained (e.g., Medicaid population would be expected to have a different distribution of debilitating mental health and vision disorders as described in Section 10.2.6), the data and code were reexamined.

9.11. Protection of human subjects

Subject information and consent

Not Applicable.

Independent Ethics Committee (IEC)/Institutional Review Board (IRB)

The protocol was amended to be a multi-database observational study (Table 2, Amendment Number 5) and was designed as an analysis of insurance claims data from large populations with health insurance. There was no active enrollment or active follow-up of patients, no data was directly collected from patients, and no medical records were abstracted. As such, approval from an IRB was not required. Only aggregated results were provided to the Sponsor or shared with other research partners.

At the time of the protocol amendment, the algorithm validation analyses for endometrial cancer and endometrial hyperplasia had already been completed in the HIRD,[24,25] thus, the validation component was removed from the protocol. This prior analysis entailed accessing Protected Health Information from medical records in order to adjudicate the primary outcomes, and IRB approval was obtained by the New England Institutional Review Board (IRB# 15-065) on 20 February 2015. A continuing review approval was obtained on 17 January 2019. Because the study relied on secondary data, informed consent was not required.

Ethical conduct of the study

The study was conducted in accordance with legal and regulatory requirements, as well as with scientific purpose, value and rigor and followed generally accepted research practices described in Guidelines for Good Pharmacoepidemiology Practices issued by the International Society for Pharmacoepidemiology, Good Epidemiological Practice guidelines issued by the International Epidemiological Association, FDA Guidance for Industry: Good Pharmacovigilance and Pharmacoepidemiologic Assessment, FDA Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting of Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets.

10. RESULTS

10.1. Participants

Of the 92,422,425 female patients with a health plan enrollment during the study period (Section 9.2), 298,450 eligible (Section 9.1) patients were identified across all five databases and all three cohorts: 18,445 new and prevalent users of CE/BZA and 280,005 new and prevalent users of E+P HT (**Appendices 8a-e, Table 1**). Regardless of other study eligibility, a total of 28,837 of 92,422,425 (0.03%) women had at least one pharmacy claim for CE/BZA during the study period (**Appendices 8a-e, Table 1**).

After PS matching, 73,294 patients were included: 18,128 CE/BZA users and 55,166 E+P HT users (**Appendices 8a-e, Table 1**). The number of PS-matched patients overall and by study database are presented in Table 5. The PS matching resulted in the exclusion of 317 of 18,445 (1.7%) CE/BZA users and 224,839 of 280,005 (80%) E+P HT users. The PS matching was not possible for Cohort 3 (prior unopposed estrogen) in three databases (Healthagen, HIRD, and MarketScan Medicaid) due to the low number of eligible patients. For these three databases for Cohort 3 (prior unopposed estrogen), unmatched (crude) data are presented. The total analytic population across the five databases included 75,455 patients: 18,417 CE/BZA users and 57,038 E+P HT users. The number of patients included in the analytic population, overall and by study database, are presented in Table 5.

The vast majority of women were commercially insured individuals, as there were only 809 women who qualified for the study from the MarketScan Medicaid database. This suggests particularly low uptake of CE/BZA in the Medicaid population, given that MarketScan Medicaid database includes over 10 million eligible women during the study period.

All included patients were classified according to their first dispensing for a study drug after at least one year of continuous enrollment.

Table 5. Number of patients in analytic population after PS-matching,* overall and by study database

	Overall	Healthagen	HIRD	MarketScan CCAE-MDCR	MarketScan Medicaid	ORD
Cohort 1	44,414	1,735	10,924	23,040	450	8,265
Cohort 2	27,451	389	7,472	13,648	82	5,860
Cohort 3	3,590	215*	1,669*	1,072	277*	357
Total	**75,455**	**2,339**	**20,065**	**37,760**	**809**	**14,482**

Abbreviations: CCAE-MDCR, commercial claims and encounters-Medicare; HIRD, HealthCore Integrated Research Database; ORD, Optum Research Database; PS, propensity score.
Cohort 1 = New users; Cohort 2 = Prevalent users; Cohort 3 = Prior unopposed estrogen.
*No PS matching for this cohort due to insufficient sample size. Crude data presented.

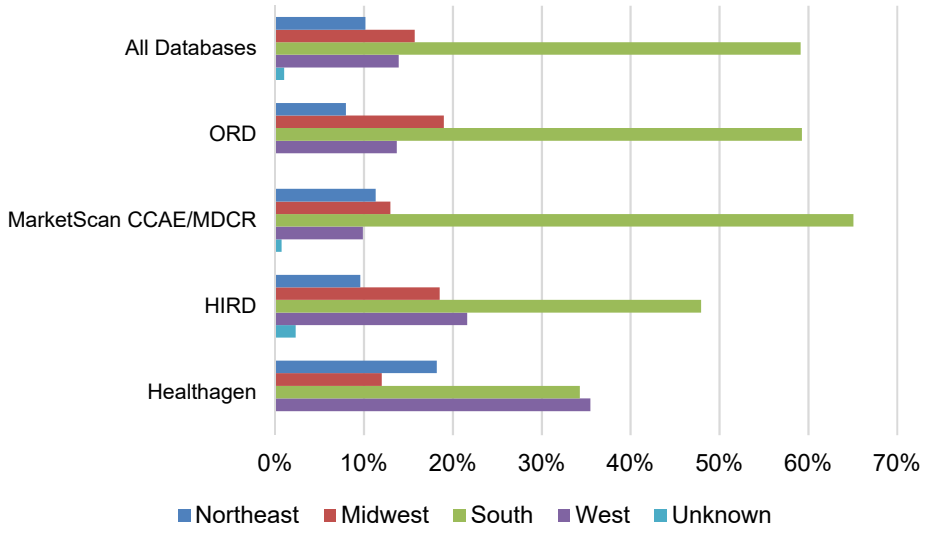
10.2. Descriptive data

Region, age, index medication, hysterectomy at follow-up, and baseline clinical covariates, overall and by database are described below.

10.2.1. Region

Data on geographic distribution of included patients, overall and by database, are presented in Figure 4. Region data were not available in the Medicaid database. Across the four databases with available data for region, most (59%) patients were located in the US South, 16% in the Midwest, 14% in the West, and 10% in the Northeastern United States. Notably, Healthagen patients were more often located in the West (35%) relative to the other databases (10-22%).

Figure 4. Geographic distribution of analytic population with available region variable, overall and by database*

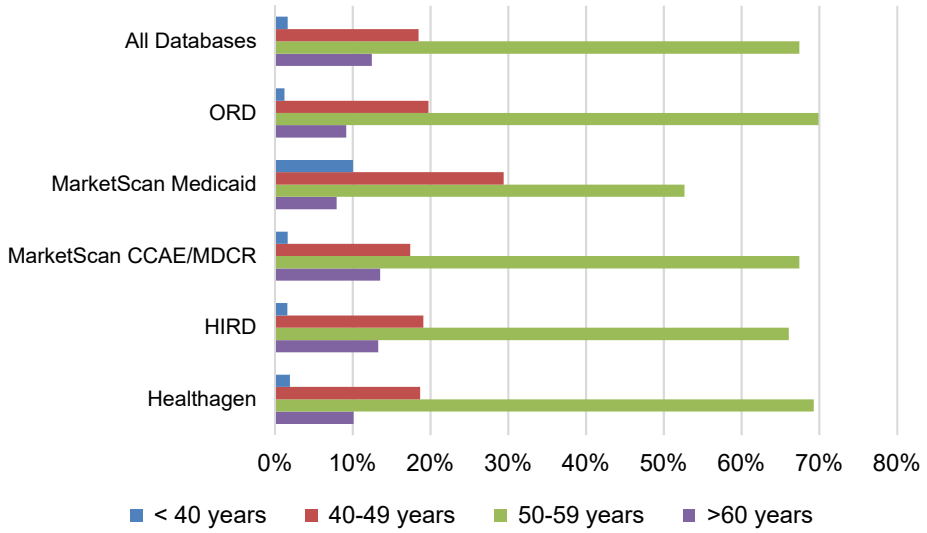


Definitions/Abbreviations: CCAE/MDCR, commercial claims and encounters-Medicare; HIRD, HealthCore Integrated Research Database; N, number; ORD, Optum Research Database.
*Excludes MarketScan Medicaid database, for which region data were not available

10.2.2. Age

Data on age (at index date) distribution of included patients, overall and by database, are presented in Figure 5. Across the five databases, most patients (67%) were 50-59 years in age, and 18% were 40-49 years in age, 12% were at least 60 years in age, and 2% were under 40 years in age. The age distribution of the Medicaid population was shifted downward, reflecting a younger population. For example, 10% of included Medicaid patients were under age 40 years while only 1-2% of included patients from the other four databases were under 40 years in age.

Figure 5. Age distribution of analytic population (n=75,455), overall and by database



Definitions/Abbreviations: CCAE/MDCR, commercial claims and encounters-Medicare; HIRD, HealthCore Integrated Research Database; N, number; ORD, Optum Research Database.

10.2.3. Index medication

Data on index medication of included patients across all five databases are presented in Table 6. Patients were classified according to their first dispensing for a study drug after at least one year of continuous enrollment. Across the five databases, 24% of patients in the analytic population (Table 5) were classified as CE/BZA users on index, and 76% as E+P HT users on index. Conjugated estrogens/medroxyprogesterone acetate tablets (Prempro and Premphase) were the most common index medication overall and among E+P HT patients (27% of analytic population). Oral estradiol/norethindrone acetate (Activella) was the index medication for 24% and transdermal estradiol/norethindrone acetate (Combipatch) for 11%. None of the included patients used estradiol-estriol-progesterone micronized cream ("Bi-Est") as their index medication (Section 9.4.1).

Table 6. US FDA approved combination hormone therapy products available by prescription in the US, by estrogen and progestin type, and distribution of index medication in analytic population across five databases[16]

Commercial Name	Estrogen Type	Progestin Type	Route	Initial US Approval*	N (%)†
Duavee	CE	N/A‡	Pill	2013	18,417 (24)
All E+P HT	--	--	--	--	57,049 (76)
Prempro	CE	Medroxyprogesterone	Pill	1995	20,500 (27)
Activella	Estradiol	Noreth acetate	Pill	1998	18,486 (24)
Combipatch	Estradiol	Noreth. acetate	Patch	1998	8,649 (11)
Femhrt	EE	Noreth. acetate	Pill	1999	5,299 (7)

Table 6. US FDA approved combination hormone therapy products available by prescription in the US, by estrogen and progestin type, and distribution of index medication in analytic population across five databases[16]

Commercial Name	Estrogen Type	Progestin Type	Route	Initial US Approval*	N (%)†
Angeliq	Estradiol	Drospirenone	Pill	2005	729 (1)
Climara Pro	Estradiol	Levonorgestrel	Patch	2003	3,298 (4)
Prefest	Estradiol	Norgestimate	Pill	1999	88 (0)

Abbreviations: CE, conjugated estrogens; EE , Ethinyl estradiol; E+P HT, Estrogen and progestin hormone therapy; FDA, Food and Drug Administration; MP, Medroxyprogesterone; N, number; N/A, not applicable; Noreth., norethindrone; US, United States.

*Source: https://www.accessdata.fda.gov/scripts/cder/daf/

†Patients in the E+P HT group could have more than one E+P HT type as their index medication. This was the case for n=11 patients (0.01%).

‡Bazedoxifene instead of progestin

Data on index medication of included patients stratified by database are presented in Table 7. Index CE/BZA use was less common in the Medicaid analytic population relative to the other four databases (16% vs. 22-26%) and conjugated estrogens/medroxyprogesterone acetate tablets (Prempro and Premphase) more common (58% vs. 25-28%). Index use of oral estradiol/norethindrone acetate (Activella) was also less common in the Medicaid analytic population relative to the other four databases (10% vs. 24-27%).

Table 7. Distribution of index medication in analytic population, by study database*

Commercial Name	Estrogen Type	Progestin Type	Healthagen	HIRD	CCAE/ MDCR	MDCD	ORD
Duavee	CE	N/A†	22%	24%	26%	16%	22%
Prempro	CE	MP	28%	28%	27%	58%	25%
Activella	Estradiol	Noreth. acetate	24%	24%	24%	10%	27%
Combipatch	Estradiol	Noreth. acetate	12%	12%	11%	7%	12%
Femhrt	EE	Noreth. acetate	7%	6%	7%	7%	7%
Angeliq	Estradiol	Drospirenone	1%	1%	1%	1%	1%
Climara Pro	Estradiol	Levonorgestrel	6%	4%	4%	2%	5%
Prefest	Estradiol	Norgestimate	0%	0%	0%	0%	0%

Abbreviations: CCAE/MDCR, commercial claims and encounters-Medicare; CE, conjugated estrogens; EE , Ethinyl estradiol; E+P HT, Estrogen and progestin hormone therapy; HIRD, HealthCore Integrated Research Database; MP, Medroxyprogesterone; N/A, not applicable; Noreth., norethindrone; ORD, Optum Research Database.

*Patients in the E+P HT group could have more than one E+P HT type as their index medication. As a result, percentages add up to >100%.

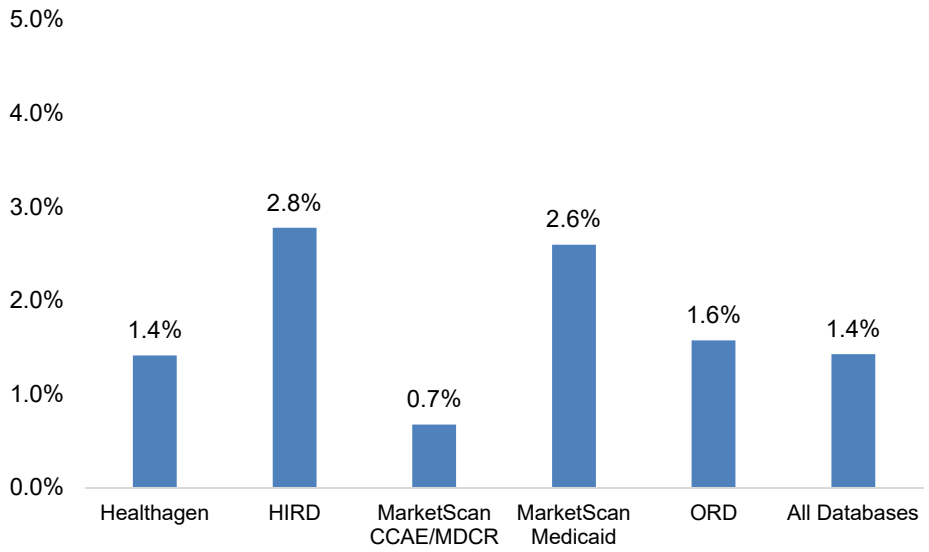
†Bazedoxifene instead of progestin.

10.2.4. Hysterectomy

Data on censoring by hysterectomy during follow-up, overall and by database, are presented in Figure 6. Data shown correspond to the Endometrial Cancer cohort analyses for all three cohorts. Across the five databases, 1.4% of included patients were censored due to hysterectomy in the Endometrial Cancer analyses. Censoring for hysterectomy among

included patients for the Endometrial Cancer analysis was notably higher in the HIRD (2.8%) and Medicaid (2.6%) relative to the other databases, and lowest in the MarketScan CCAE & Medicare Supplemental database (0.7%;Figure 6). Censoring by hysterectomy in the Endometrial Cancer analysis was similar between included CE/BZA users (1.37%) and E+P HT users (1.45%).

Figure 6. Percentage of analytic population censored for hysterectomy in the Endometrial Cancer analysis, overall and by database



Definitions/Abbreviations: CCAE/MDCR, commercial claims and encounters-Medicare; HIRD, HealthCore Integrated Research Database; N, number; ORD, Optum Research Database.

10.2.5. Length of baseline period

Baseline period, i.e., period of continuous enrollment prior to index, varied by database and cohort, with cohort-specific medians ranging from 13-41 months (Table 8): 31-41 months for Healthagen, 15.9-40.9 for HIRD, 15.2-29.8 months for MarketScan CCAE & Medicare Supplemental, 14.0-29.3 months for MarketScan Medicaid, and 13-38 months for ORD. In all databases except for Healthagen, Cohort 2 (prevalent users) tended to have less baseline time than the other cohorts (13-16.2 months vs. 22.6-40.9 months). For Healthagen, median baseline was 40-41 months in Cohort 2 (prevalent users) and 31-34 in Cohorts 1 and 3. Across all study databases, for Cohort 1, median continuous enrollment prior to index date was 22.6 to 38 months, and individual continuous enrollment prior to index date ranged from 12 to 164 months (**Appendices 8a-e, Table 3**).

Table 8. Median continuous enrollment time prior to index date, by study database, in months

Database	Median Baseline (months)
All databases/cohorts (n=15)*	**13.0-41.0**
Healthagen	31.0-41.0
HIRD	15.9-40.9
MarketScan CCAE & Medicare Supplemental	15.2-29.8
MarketScan Medicaid	14.0-29.3
ORD	13.0-38.0

Abbreviations: CCAE, commercial claims and encounters; HIRD, HealthCore Integrated Research Database; N, number; ORD, Optum Research Database.
*Each median corresponds to a specific cohort; e.g., Healthagen, Cohort 2. There are 15 medians, representing 5 databases with 3 cohorts each.

10.2.6. Baseline clinical covariates

Across databases and cohorts, the most common diagnoses among included patients were medical exam/evaluation (61.5-97.7%; **Appendices 8a-e, Tables 2a-c**), "other screening for suspected conditions" (57.6-94.8%; **Appendices 8a-e, Tables 2a-c**), and menopausal disorders (37.4-83.0%; **Appendices 8a-e, 2a-c**). Other common diagnoses included other connective tissue disease (23.5-81.8%; **Appendices 8a-e, 2a-c**), spondylosis/disc disorders/other back problems (20.4-74.4%; **Appendices 8a-e, 2a-c**). Disorders of hyperlipidemia (18.0-63.6%; **Appendices 8a-e, Table 3**) and hypertension (19.2-81.8%; **Appendices 8a-e, Table 3**) were also prevalent, as were thyroid disorders (18.2-42.9%; **Appendices 8a-e, Tables 2a-c**). Patients in Cohort 3 (prior unopposed estrogen) and Medicaid patients generally showed greater percentages for baseline comorbidities (**Appendices 8a-e, Tables 2a-c**).

Medicaid patients exhibited a higher prevalence of mood disorders (39.0-81.8% vs. 10.5-19.6% in CCAE/Medicare and ORD; not present in Top 25 diagnoses for other databases; **Appendices 8a-e, Tables 2a-c**), anxiety disorders (36.6-63.6% vs. 12.2-40.0% in HIRD, CCAE/Medicare, and ORD; not present in Top 25 for Healthagen), and blindness and vision defects (39.0-72.7% vs. 27.3-40.5%; not present in Top 25 for HIRD or ORD; **Appendices 8a-e, Tables 2a-c**).

Baseline mammography was more common in the Healthagen, HIRD, and MarketScan CCAE & Medicare Supplemental cohorts (75.3-87.9%) than in the MarketScan Medicaid or ORD cohorts (31.9-55.6%).

10.3. Outcome data

For each outcome, fewer than 1% of the analytic population experienced an outcome prior to censoring, with the exception of the secondary outcome of "any cancer (other than NMSC)," which affected slightly over 1% across cohorts. For the main outcomes, a total of 39 endometrial cancer and 48 endometrial hyperplasia cases occurred in the 44,414 PS-matched

patients in Cohort 1 (new users). In the 27,451 PS-matched members of Cohort 2 (prevalent users), there were 14 and 32 cases of endometrial cancer and endometrial hyperplasia, respectively. In the 3,590 members of analytic Cohort 3 (prior unopposed estrogen), there were 3 and 7 cases of endometrial cancer and endometrial hyperplasia, respectively. The most frequent secondary outcomes in the analytic population were any cancer excluding NMSC (n=618 in Cohort 1, n=451 in Cohort 2, n=44 in Cohort 3) and breast cancer (n=286 in Cohort 1, n=228 in Cohort 2, and n=18 in Cohort 3). The least frequent secondary outcome in the analytic population was renal cancer and renal adenoma (n=9 in Cohort 1, n=8 in Cohort 2, n=1 in Cohort 3).

10.4. Main results

Main findings are presented in Figure 7 and Figure 8 and in **Appendix 8f, Table 11**.

10.4.1. Primary analyses – New Users (Cohort 1)

10.4.1.1. Overall, pooled and unpooled

Endometrial Cancer

Across all five databases in the Cohort 1 (new users) PS-matched population, a total of 39 endometrial cancer cases occurred in 82,458 person-years of follow-up (IRR_{pooled}=1.50; 95% CI: 0.79, 2.88;Figure 7; and HR_{pooled}=1.50; 95% CI: 0.75, 2.98;Figure 8): 12 cases among CE/BZA users with 19,704 person-years of follow-up (IR_{pooled}=5.20 per 10,000 person-years; 95% CI: 2.02, 8.38) and 27 cases among E+P HT users with 62,754 person-years of follow-up (IR_{pooled}=3.60 per 10,000 person-years; 95% CI: 1.13, 6.07). The pooled HR for endometrial cancer represents HIRD and MarketScan CCAE & Medicare Supplemental data. The remaining three databases did not produce cases of endometrial cancer in both the CE/BZA user and the E+P HT user groups for the PS-matched new user cohort (Cohort 1). All databases were included in the pooled IRR calculation (**Appendix 8g**).

Endometrial Hyperplasia

In the Cohort 1 PS-matched population, a total of 48 endometrial hyperplasia cases occurred in 38,770 person-years of follow-up (IRR_{pooled}=1.69; 95% CI:0.51, 5.61,Figure 8; and HR_{pooled}=1.79; 95% CI: 0.43, 7.54; Figure 7)14 cases among CE/BZA users with 9,689 person-years of follow-up (IR_{pooled}=11.00 per 10,000 person-years; 95% CI: 1.84, 20.17) and 34 cases among E+P HT users with 29,081 person-years of follow-up (IR_{pooled}=10.60 per 10,000 person-years; 95% CI: 6.13, 15.07). Similar to the endometrial cancer outcome, the pooled HR for endometrial hyperplasia represents HIRD and MarketScan CCAE & Medicare Supplemental data. The remaining three databases did not produce cases of endometrial hyperplasia in both the CE/BZA user and the E+P HT user groups for the PS-matched new user cohort (Cohort 1). All databases were included in the pooled IRR calculation (**Appendix 8g**).

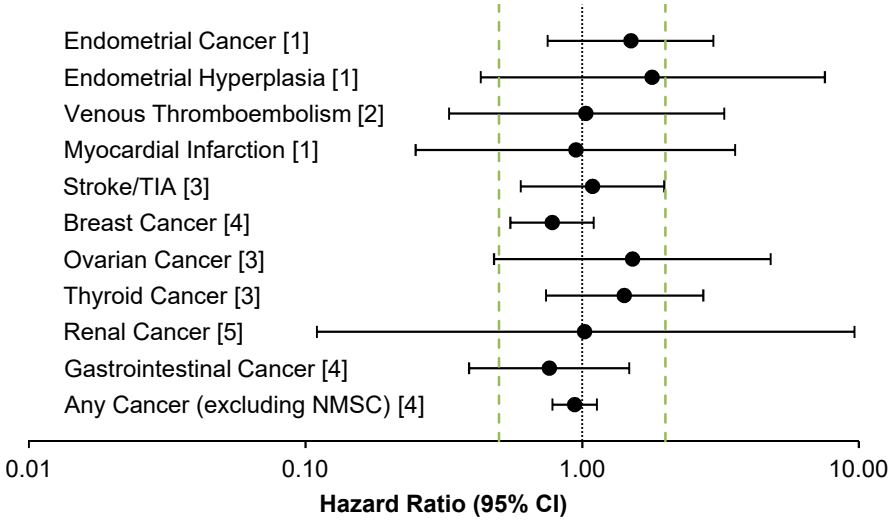
Other Cancers

The risks of most other forms of cancer were largely similar comparing the CE/BZA and E+P HT new users in the Cohort 1 PS-matched population, including any form of cancer other than NMSC (n=618; IRR_{pooled}=0.93; 95% CI: 0.77, 1.13; Figure 8; HR_{pooled}=0.94; 95% CI: 0.78, 1.13; Figure 7). This was also true for specific cancers of interest, with pooled IRR estimates ranging from 0.79 to 1.89. There were 286 breast cancer cases (IRR_{pooled}=0.79; 95% CI: 0.58, 1.05; Figure 8; HR_{pooled}=0.78; 95% CI: 0.55, 1.10; Figure 7), 15 ovarian cancer cases, (IRR_{pooled}=1.89; 95% CI: 0.65, 5.47; Figure 8; HR_{pooled}=1.52; 95% CI: 0.48, 4.80; Figure 7), 43 thyroid cancer cases (IRR_{pooled}=1.50; 95% CI: 0.79, 2.85; Figure 8; HR_{pooled}=1.42; 95% CI: 0.74, 2.74; Figure 7), nine renal cancer/adenoma cases (IRR_{pooled}=1.07; 95% CI: 0.29, 4.02; Figure 8; HR_{pooled}=1.02; 95% CI: 0.11, 9.65; Figure 7), 59 gastrointestinal cancer cases (IRR_{pooled}=0.79; 95% CI: 0.41, 1.51; Figure 8; HR_{pooled}=0.76; 95% CI: 0.39, 1.48; Figure 7). Pooled HRs for the cancer outcomes represent databases with non-zero case values for each treatment group, as specified in **Appendix 8f, Table 11**. All databases were included in the pooled IRR calculations (**Appendix 8g**).

Cardiovascular Outcomes

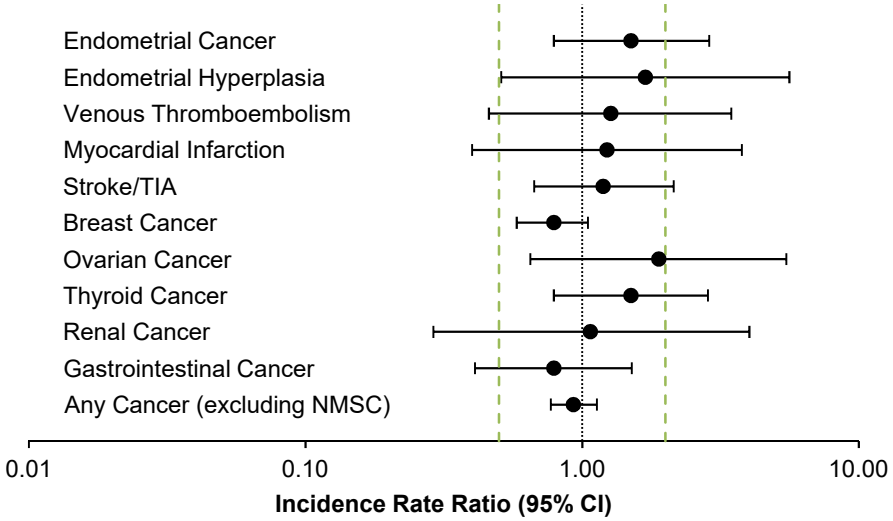
The three secondary acute outcomes were also similar comparing the CE/BZA and E+P HT new users, with pooled IRR estimates ranging from 1.19 to 1.27 (Figure 8 and Figure 7). There were 17 VTEs (IRR_{pooled}=1.27; 95% CI: 0.46, 3.46; HR_{pooled}=1.03; 95% CI: 0.33, 3.26), 14 MIs (IRR_{pooled}=1.23; 95% CI: 0.40, 3.78; HR_{pooled}=0.95; 95% CI: 0.25, 3.57), and 56 strokes/TIAs (IRR_{pooled}=1.19; 95% CI: 0.67, 2.14; HR_{pooled}=1.09; 95% CI: 0.60, 1.98). Pooled HRs for the secondary acute outcomes represent databases with non-zero case values for each treatment group, as specified in **Appendix 8f, Table 11**. All databases were included in the pooled IRR calculations (**Appendix 8g**).

Figure 7. Pooled Hazard Ratios for two main outcomes and nine secondary outcomes across five study databases [1-5], comparing CE/BZA and E+P HT new users (Cohort 1).



Abbreviations: CCAE, claims and commercial encounters; CE/BZA, Conjugated estrogens/bazedoxifene; CI, confidence interval; E+P HT, Estrogen and progestin hormone therapy; HIRD, HealthCore Integrated Research Database; HR, hazard ratio; NMSC, Non-melanoma skin cancer; ORD, Optum Research Database; TIA, transient ischemic attack.
[1] Pooled HR represents HIRD and MarketScan CCAE & Medicare Supplemental data due to zero counts in the other databases.
[2] Pooled HR represents MarketScan CCAE & Medicare Supplemental and ORD data due to zero counts in the other databases.
[3] Pooled HR represents HIRD, MarketScan CCAE & Medicare Supplemental, and ORD data due to zero counts in the other databases.
[4] Pooled HR represents Healthagen, HIRD, MarketScan CCAE & Medicare Supplemental, and ORD data due to zero counts in the other databases.
[5] Pooled HR represents HIRD only due to zero counts in the other databases.

Figure 8. Pooled Incidence Rate Ratios for two main outcomes and nine secondary outcomes across five study databases, comparing CE/BZA and E+P HT new users (Cohort 1).



Abbreviations: CE/BZA, Conjugated estrogens/bazedoxifene; CI, confidence interval; E+P HT, Estrogen and progestin hormone therapy; NMSC, Non-melanoma skin cancer; TIA, transient ischemic attack.

10.4.1.2. Stratified

Within each of the five databases, endometrial cancer and endometrial hyperplasia analyses in new users (Cohort 1) were stratified by age group, treatment duration, timing of use, use of topical hormonal treatments, and prior osteoporosis (**Appendices 8a-e, Tables 8b & 8c**). Given the rarity of both endometrial hyperplasia and cancer in these study populations, stratification by these variables produced few estimates for endometrial cancer (**Appendices 8a-e, Table 8b**) or hyperplasia (**Appendices 8a-e, Table 8c**) due to zero case counts for most strata. For those strata with non-zero case counts, precision was poor due to limited data.

For the main and stratified analyses, the Healthagen, MarketScan Medicaid, and ORD databases did not report HRs for endometrial cancer or hyperplasia due to zero case counts. Stratified estimates were available for the HIRD and MarketScan CCAE & Medicare Supplemental databases, and while imprecise (i.e., CIs for point estimates were wide and overlapped the null value), results were largely similar between groups in these stratified analyses. One exception was noted in the MarketScan CCAE & Medicare Supplemental database where the HRs for endometrial cancer and endometrial hyperplasia differed in magnitude and direction – though with extremely wide CIs due to the low case counts. For example, while the HR for endometrial cancer in the main analysis was 1.35, it was 0.70 (95% CI: 0.08, 5.98) for the age ≥60 stratum and 2.12 (95% CI: 0.35, 12.70) for the long duration of use stratum.

Results were also largely similar in individual databases when there were sufficient cases available to perform an analysis. However, the HIRD data generated a high magnitude HR

for endometrial hyperplasia (HR_{HIRD}= 3.77; 95% CI: 1.31, 10.90; **Appendix 8b, Table 8a**). While imprecise due to a limited number of cases (<11), this association held for the following strata: age 50-59 years (HR=4.26; 95% CI: 1.20, 15.10), all duration of use strata (HRs from 2.02 to 4.08 and 95% CI ranging from 1.08 to 10.90), no topical hormone treatments (HR= 4.29; 95% CI: 1.36, 13.50), and no prior osteoporosis (HR= 5.03; 95% CI: 1.47, 17.20). All HIRD endometrial hyperplasia HRs were elevated except for in the short duration of use stratum (HR= 0.88, 95% CI: 0.18, 4.36). The only other database with sufficient cases to report an estimate of the HR for short duration of use was the MarketScan CCAE & Medicare Supplemental database, which also reported a HR for CE/BZA vs. E+P HT use in the protective direction in this group (HR=0.57; 95% CI: 0.13, 2.62).

10.4.2. Secondary analyses – prevalent users (Cohort 2, Prior CE/BZA or E+P HT Use; Cohort 3, Prior Unopposed Estrogen Use)

10.4.2.1. Overall, pooled and unpooled

Comparative analyses were not conducted for Cohorts 2 and 3 (prevalent users of study drugs or unopposed estrogen) due to concerns about prevalent user bias.[17] Instead, IRs were calculated and reported for each treatment group for each cohort. Pooled incidence rates for all outcomes for Cohorts 2 and 3 are displayed in **Appendix 8f, Table 11**.

Pooled incidence rates for endometrial cancer ranged from 2.10 per 10,000 person-years in E+P HT users with prior use of a study drug (Cohort 2) to 13.07 per 10,000 person-years in CE/BZA users with prior use of unopposed estrogen (Cohort 3). Pooled incidence rates for endometrial hyperplasia ranged from 7.75 per 10,000 person-years in CE/BZA users with prior use of a study drug (Cohort 2) to 34.78 per 10,000 person-years in CE/BZA users with prior use of unopposed estrogen (Cohort 3). Secondary outcome rates varied widely (**Appendix 8f, Table 11**) and are summarized in Table 9.

Table 9. Ranges of incidence rates for each cohort from each database (n=36*) for secondary outcomes in Cohorts 2 and 3

Outcome	Rate Range
Venous thromboembolism	4.77 to 24.53
Myocardial infarction and/or sudden death	2.62 to 24.11
Stroke (including transient ischemic attack)	13.38 to 34.66
Breast cancer	30.01 to 54.68
Ovarian cancer	2.05 to 8.53
Thyroid cancer	2.58 to 17.79
Renal cancer/Renal Adenoma	1.17 to 8.53
Gastrointestinal cancer	8.36 to 10.16
Cancer (all types excluding NMSC)	58.77 to 97.01

Abbreviations: N, number; NMSC, non-melanoma skin cancer.
*There are four incidence rates for each of the nine outcomes – i.e., one for each of treatment groups for each of the two cohorts

10.4.2.2. Stratified

Within each of the five databases, endometrial cancer and endometrial hyperplasia analyses in prevalent users (Cohorts 2 and 3) were stratified by age group, treatment duration, timing of use, use of topical hormonal treatments, and prior osteoporosis (**Appendices 8a-e, Tables 8b & 8c**).

Endometrial Cancer

In three of the five databases (Healthagen, MarketScan Medicaid, and ORD), there were 0 cases of endometrial cancer among prevalent users, resulting in stratified IRs of 0. Stratified IRs for endometrial cancer were available from the HIRD and the MarketScan CCAE & Medicare Supplemental databases:

Prior users of CE/BZA or E+P HT (Cohort 2). In CE/BZA users with prior use of CE/BZA or E+P HT (Cohort 2), IRs were highest in the following stratum in both databases: age ≥60 years, and no prior osteoporosis. However, these findings are to be interpreted with caution given the low number of cases in each of these strata. Among the E+P HT users in Cohort 2, IRs were highest in the following stratum in both databases: age ≥60 years, duration of use ≤1 year (short), no prior topical hormonal treatments, and no prior osteoporosis.

Prior users of unopposed estrogen (Cohort 3). In CE/BZA users with prior use of unopposed estrogen (Cohort 3), IRs were highest in the following stratum in the MarketScan CCAE & Medicare Supplemental database: age ≥60 years, duration of use ≤1 year (short), current use, no prior topical hormonal treatments, and no prior osteoporosis. These findings are to be interpreted with caution given the low number of cases (n=1 or 0) in each of these strata. Stratified IRs were not available for CE/BZA users in Cohort 3 in the HIRD due to 0 reported cases for this population. In E+P HT users with prior use of unopposed estrogen (Cohort 3), IRs were highest in the following stratum in both databases: ever use, and no prior osteoporosis. Again, these findings are to be interpreted with caution given the low number of cases in each stratum.

Consistent findings that IRs were higher in the no prior topical hormonal treatments and no prior osteoporosis strata may be attributable to the fact that these strata consistently contained higher numbers of person-time (i.e., more follow-up could translate to more outcomes that were captured). In contrast, age ≥60 years consistently showed higher IRs for endometrial cancer while age 50-59 consistently contained the most person-time.

Endometrial Hyperplasia

In the Healthagen and MarketScan Medicaid databases, there were 0 cases of endometrial hyperplasia among prevalent users, resulting in stratified IRs of 0. Stratified IRs for endometrial hyperplasia were available for three of the five databases (HIRD, MarketScan CCAE & Medicare Supplemental, ORD):

Prior users of CE/BZA or E+P HT (Cohort 2). In the MarketScan CCAE & Medicare Supplemental database, there were 0 cases of endometrial hyperplasia in the CE/BZA group of Cohort 2, resulting in stratified IRs of 0. In the CE/BZA group of Cohort 2 in the HIRD and ORD, IRs were highest for age ≥60 years, and no prior topical hormonal treatments. In the E+P HT users in Cohort 2, IRs were highest in age ≥60 years in all three databases. Results were inconsistent for other stratification variables for this cohort across databases.

Prior users of unopposed estrogen (Cohort 3). Among the CE/BZA users in Cohort 3 in all three databases, IRs were highest among those with no prior topical hormonal treatments, or no prior osteoporosis. Among the E+P users in Cohort 3 in all three databases, IRs were highest in age ≥60 years, duration of use ≤1 year (short), no prior topical hormonal treatments, and no prior osteoporosis.

Consistent findings that IRs were higher in the no prior topical hormonal treatments or no prior osteoporosis strata may be attributable to the fact that these strata consistently contained higher numbers of person-time. In contrast, age ≥60 years consistently showed higher IRs for endometrial cancer while age 50-59 consistently contained the most person-time.

10.5. Other analyses

10.5.1. Sensitivity analyses of main results primary analyses

10.5.1.1. Fixed effects pooling

Pooled HRs (reported in Section 10.4) were based upon findings from a random effects model approach. All of HRs were re-run using a fixed effects approach. For most outcomes, fixed effects models produced the same HRs and 95% CIs as the random effects models. For endometrial hyperplasia, the fixed effects model produced a minute attenuation of the magnitude of the HR, with a narrowing of the 95% CI width (HR_{random}=1.79; 95% CI: 0.43, 7.54 to HR_{fixed}=1.74; 95% CI: 0.84, 3.59). The breast cancer findings were similarly affected, with a slight narrowing of the 95% CI width (HR_{random}= 0.78; 95% CI: 0.55, 1.10 to HR_{fixed}=0.78; 95% CI: 0.58, 1.06).

Changes to Censoring Criteria. When censoring due to a change in treatment during follow-up was removed among new users (Cohort 1), estimates did not differ appreciably in any database. In the two largest databases, the HIRD and the MarketScan CCAE & Medicare Supplemental database, the point estimates for the HR of this sensitivity analysis for endometrial cancer and endometrial hyperplasia were all within 0.1 of the main estimate. The endometrial cancer estimates were HR=1.65 (95% CI: 0.55, 4.93) and HR=1.46 (95% CI: 0.63, 3.42), respectively. Endometrial hyperplasia estimates were HR=3.72 (95% CI: 1.29,

10.70) in the HIRD and HR=0.86 (95% CI: 0.32, 2.34) in the MarketScan CCAE & Medicare Supplemental database. Breast cancer estimates remained lower among CE/BZA users than in E+P HT users in most of the databases, and were HR= 0.78 (95% CI: 0.47, 1.32) in the HIRD, HR= 0.84 (95% CI: 0.57, 1.25) in the MarketScan CCAE & Medicare Supplemental database, and HR=0.37 (95% CI: 0.15, 0.94) in the ORD, but HR was 2.00 (95% CI: 0.18, 22.05) in the Healthagen database (**Appendices 8a-e, Table 9**).

Similarly, when censoring due to a cancer diagnosis during follow-up for non-cancer outcomes was removed among new users (Cohort 1), estimates did not differ appreciably in any database (**Appendices 8a-e, Table 9**).

Changes to Time Period for Primary Outcomes and Acute, Secondary Outcomes. In the analysis among new users (Cohort 1) restricting to the first new-use treatment episode, endometrial hyperplasia findings increased in magnitude to HR= 6.37 (95% CI: 1.59, 25.50) in the HIRD and to HR=1.93 (95% CI: 0.63, 5.91) in the MarketScan CCAE & Medicare Supplemental database (**Appendices 8a-e, Table 9)**. Some of the results for the other outcomes were also changed, often increasing in magnitude in the same direction of the original estimate. Notably, in the ORD, the stroke/TIA estimate changed from HR=1.26 (95% CI: 0.32, 4.91) in the main analysis to HR=0.48 (95% CI: 0.06, 4.05) when restricted to the first new use treatment episode.

Changes to Cancer Definition. When the definition of "any cancer" was expanded to include NMSC (**Appendices 8a-e, Table 9**) for new users (Cohort 1), estimates differed only minutely and remained slightly on the protective side of the null (HR=1), with the exception of the Healthagen data (HR=1.40, 95% CI: 0.62, 3.15).

Changes to Time Period for Acute, Secondary Outcomes. When the time period for assessing acute secondary outcomes among new users (Cohort 1) was expanded to include recent use periods (in addition to person-time from current use periods), findings did not differ appreciably for any database (**Appendices 8a-e, Table 9**). Of note, the HIRD findings for VTE using current use time alone did not have enough cases to calculate a hazard ratio. After expanding the follow-up period to include recent use time, additional cases were acquired and the HR was 0.97 (95% CI: 0.10, 9.28). Simultaneously, the HIRD HRs for MI and stroke/TIA moved further from the null, going from HR=0.48 (95% CI: 0.06, 4.02) to HR=0.29 (95% CI: 0.04, 2.23) for MI and from HR=0.47 (95% CI: 0.06, 3.89) to HR=0.36 (95% CI: 0.04, 2.85) for stroke/TIA.

10.5.1.2. Pooled results excluding MarketScan CCAE & Medicare supplemental database

Because MarketScan CCAE & Medicare Supplemental database may include events and person-time that are also part of the HIRD, ORD, or Healthagen databases, pooled estimates were re-calculated excluding MarketScan CCAE & Medicare Supplemental data (**Appendix 8f, Table 12**).

Given that there were only cases of endometrial hyperplasia and cancer in both the CE/BZA and E+P HT new users in HIRD and MarketScan CCAE databases, hazard ratios were only able to be calculated in those databases. Thus, when pooling the endometrial hyperplasia and cancer results without MarketScan CCAE, the pooled HR results included only the HIRD results: HR=1.76 (95% CI: 0.59, 5.25) for endometrial cancer and HR=3.77 (95% CI: 1.31, 10.87) for endometrial hyperplasia.

Estimates for secondary outcomes did not follow a single pattern in how they changed upon the removal of the MarketScan CCAE & Medicare Supplemental database patients. Some estimates moved away from the null, while others moved towards the null, and others remained the same. For example, the estimates for venous thromboembolism, myocardial infarction, and ovarian cancer moved away from the null: from HR=1.03 (95% CI: 0.33, 3.26) to HR=1.56 (0.28, 8.69) for venous thromboembolism, from HR=0.95 (0.25, 3.57) to HR=0.48 (95% CI: 0.06, 3.93) for myocardial infarction, and from HR=1.52 (95% CI: 0.48, 4.80) to HR=1.94 (95% CI: 0.27, 13.97) for ovarian cancer. The gastrointestinal cancer estimate moved towards the null, from HR= 0.76 (95% CI: 0.39, 1.48) to HR=0.93 (95% CI: 0.40, 2.15) along with the findings for thyroid cancer, which changed from HR= 1.42 (95% CI: 0.74, 2.74) to HR=1.26 (95% CI: 0.31, 5.07). The remaining secondary outcomes did not change appreciably. Finally, the estimate for breast cancer changed only minutely, from HR=0.78 (95% CI: 0.55, 1.10) to HR=0.74 (95% CI: 0.38, 1.43).

10.5.1.3. Quantitative bias analysis

This study utilized previously validated, HIRD-based algorithms to identify endometrial cancer and endometrial hyperplasia.[23,24] In women who were users of E+P HT in the HIRD, the endometrial cancer algorithm had a PPV of 90.8% (95% CI 86.9 – 93.6).[24] Given the observed pooled IRR of 1.50, if PPV was 89% in one treatment group and 93% in the other, and sensitivity was 78% in one group, and 82% in the other, the corrected IRRs would range from 1.37 to 1.65 (**Appendix 8f, Table 13**). In new users of E+P HT in the HIRD, the predictive model-based endometrial hyperplasia algorithm was developed and applied to the present study to yield a PPV of 80% and a sensitivity of up to 69% (assuming a 100% sensitivity for the endometrial hyperplasia screening algorithm used to identify the validation sample).[23] Given the observed pooled IRR of 1.69, if PPV was 78% in one treatment group and 82% in the other, and sensitivity was 48% in one group, and 52% in the other, the outcome-misclassification corrected IRRs would range from 1.48 to 1.92 (**Appendix 8f, Table 13**).

These estimates were generated using a published formula derived for the correction of relative risk estimates.[29] An important feature of the formula is that whenever PPV and sensitivity are both the same between treatment groups, the corrected estimate is equal to the observed estimate, implying no bias. This approach is limited by the fact that the previous validation work[23,24] did not stratify PPV or sensitivity by treatment group, and thus the extent of difference in misclassification remains unknown. It is possible that CE/BZA users were more closely monitored for endometrial cancer or hyperplasia given the drug's novelty to both patients and clinicians, and that closer monitoring would theoretically lead to a higher sensitivity in claims in the CE/BZA users relative to the E+P users. However, for the present

findings for endometrial hyperplasia and endometrial cancer to be nullified or reversed (i.e., IRR≤1.0), each outcome's sensitivity would need to be markedly higher in the CE/BZA users than in the E+P users. For example, if the sensitivity for endometrial cancer was 100% in CE/BZA users, the sensitivity in E+P users would need to be 70% or lower in order to attenuate or reverse the IRR (IRR≤1.0). Another way that the present findings would be reversed or nullified (i.e., IRR≤1.0) is if the sensitivity was extremely low (<15%) in one or both treatment groups while also being at least slightly higher in the CE/BZA group than in the E+P HT group.

An additional limitation is that the previous validation work included E+P HT users only. The PPV in this population might not be transportable to CE/BZA users, particularly CE/BZA users who never used E+P HT, due to differences in underlying EC/EH prevalence. Finally, neither algorithm directly assessed sensitivity.

10.5.2. Quantification of loss-to-follow-up due to Medicare enrollment

Because post-menopausal women are in the age group approaching Medicare eligibility (age 65 years), there was a potential for loss-to-follow-up related to disenrollment from employer-sponsored insurance plans and enrollment in Medicare. If women enrolled in Medicare prior to developing an outcome or another censoring criterion, then that outcome would not be captured in the present study, which relies solely on non-Medicare claims. This is of particular importance for cancer, which may take longer to develop and to be detected than acute outcomes such as myocardial infarction.

The databases included in this study did not contain variables indicating member disenrollment due to transition to Medicare, and the extent of the potential issue could not be directly quantified. Instead, to quantify the potential loss-to-follow-up due to transition to Medicare, disenrollments on or around patients' 65th birthdays were counted and reported in **Appendix 8f, Table 14**, for each database and for Cohort 1 (new users) before and after PS matching.

This analysis found that censoring related to Medicare enrollment would be unlikely to have much impact on this study, as almost all the women in this study were under 65. Indeed, only 1% to 4% of patients (varied by database) in the PS matched Cohort 1 (new users) endometrial cancer analysis remained in the study at age 64 years and 11 months (age at which they would have become eligible for Medicare initiation) or 64th year of life (for MarketScan database patients, for whom no birthdate was available). Of these women, 7.7% to 67.5% disenrolled on or after the year of their 65th birthday prior to experiencing endometrial cancer or censoring event or end of study period (**Appendix 8f, Table 14**). This loss was lowest in the MarketScan Medicaid cohort (7.7%), highest in the ORD (67.5%), and was approximately 50% for Healthagen, HIRD, and MarketScan CCAE & Medicare Supplemental databases.

A separate analysis examined disenrollment in all females in each of the five databases, regardless of whether they qualified for this study. Across databases, 49.6% to 69.5% of women with enrollment in the study period (01 May 2014 to 31 August 2019) who were aged 64 years disenrolled on or after the year of their 65th birthday. It appears that Medicare-

related disenrollment is common across databases among women who reach Medicare eligibility age.

10.6. Adverse events / adverse reactions

This multi-database study included claims data that were converted to structured (i.e., coded) data solely by a computer using automated/algorithmic methods and/or data that already existed as structured data in an electronic database. In these data sources, it was not possible to link (i.e., identify a potential association between) a particular product and medical event for any individual. Thus, the minimum criteria for reporting an AE (i.e., identifiable patient, identifiable reporter, a suspect product, and event) were not available and AEs were not reportable as individual AE reports.

The protocol was amended to be a multi-database observational study (Table 2, Amendment Number 5) and was designed as an analysis of insurance claims data from large populations with health insurance. At the time of the protocol amendment, the algorithm validation analyses for endometrial cancer and endometrial hyperplasia had already been completed in the HIRD,[24,25] thus, the validation component was removed from the protocol. For this prior analysis, human review of patient-level unstructured data was required. Unstructured data referred to verbatim medical data, including text-based descriptions and visual depictions of medical information, such as medical records, images of physician notes, neurological scans, X-rays, or narrative fields. The reviewer was obligated to report adverse event(s) (AE) with explicit attribution to any Pfizer drug that appeared in the reviewed information. No reportable safety information was discovered during this validation exercise.

11. DISCUSSION

11.1. Key results

This study estimated and compared the rates of endometrial cancer and endometrial hyperplasia among postmenopausal women initiating CE/BZA and postmenopausal women initiating estrogen and progestin combination hormone therapy (E+P HT) during the first five years of CE/BZA availability in the US. In this study, which included five large databases, CE/BZA use was uncommon with only 28,837 of 92,422,425 (0.03%) women using CE/BZA during the study period. The PS-matched, comparative analyses included 10,596 CE/BZA and 33,818 E+P HT users (i.e., Cohort 1, new users).

This study found that for the comparison of CE/BZA users to E+P HT users, the ratio of incidence rates for endometrial cancer was greater than 1.00 but with a 95% CI that included 1.00 (IRR_{pooled}=1.50; 95% CI: 0.79, 2.88; HR_{pooled}=1.50; 95% CI: 0.75, 2.98). This outcome was uncommon, thus, the study had limited statistical power to estimate small or moderate magnitudes of effect. Results were largely similar when MarketScan CCAE & Medicare Supplemental patients were removed, and across sensitivity analyses addressing censoring criteria, outcome definitions, and potential overlap between databases. Endometrial hyperplasia findings were similar to those for endometrial cancer (IRR_{pooled}=1.69; 95% CI: 0.51, 5.61; HR_{pooled}=1.79; 95% CI: 0.43, 7.54) but increased in magnitude upon removal of the MarketScan CCAE & Medicare Supplemental patients (IRR_{pooled}=2.35; 95% CI: 0.47,

11.83; HR_{pooled}=3.77; 95% CI: 1.31, 10.87). The pooled HRs, however, represented only the HIRD and/or MarketScan CCAE & Medicare Supplemental data given that the other databases had 0 cases in either the CE/BZA or E+P HT groups. Thus, when MarketScan CCAE & Medicare Supplemental patients were removed, only the HIRD data alone were represented. The pooled IRRs included all 5 databases and were 1.50 (95% CI: 0.79, 2.88) for endometrial cancer and 1.69 (95% CI:0.51, 5.61) for endometrial hyperplasia.

The inclusion of progestin in E+P HT is intended to protect patients against endometrial cancer (when taken instead of unopposed estrogen). A meta-analysis pooling results from 37 studies found that the risk of endometrial cancer increased about two-fold in women who had ever used unopposed estrogen therapy compared to non-users, but varied substantially by duration, dose, and type of estrogen used.[30] In contrast, their meta-analysis found that cohort studies suggested a decrease in risk of endometrial cancer in E+P HT vs. non-users (relative risk [RR]=0.4), although case-control studies did suggest an increase in risk vs. non-users (RR=1.8).[30] CE/BZA includes bazedoxifene instead of progestin. Bazedoxifene is a SERM, which prevents cellular uptake of estrogen and therefore also has an antagonizing, oppositional effect on estrogen, and could plausibly confer protection against endometrial cancer in women taking exogenous estrogen. The present study found that for the comparison of the CE/BZA group to the E+P HT group, the ratios of incidence rates for endometrial cancer and endometrial hyperplasia were both greater than 1.00, though the magnitude of the effect was small and 95% CIs were wide, overlapping the null value. Whether bazedoxifene is more or less protective than progestin against endometrial cancer remains relatively uncertain due to low event counts and limited statistical power in this study.

Additionally, this study estimated and compared the rates of three cardiovascular and six cancer outcomes among postmenopausal women initiating CE/BZA and postmenopausal women initiating E+P HT. This study found a lower rate of breast cancer in CE/BZA users relative to E+P HT users (IRR_{pooled}=0.79; 95% CI: 0.58, 1.05 and HR_{pooled}=0.78; 95% CI: 0.55,1.10). This finding held, though with less precision, when the MarketScan CCAE & Medicare Supplemental patients were removed (IRR_{pooled}= 0.73; 95% CI: 0.45, 1.20 and HR_{pooled}=0.74; 95% CI: 0.38, 1.43). This finding is notable since women who are concerned with breast cancer risk may be channeled to CE/BZA instead of E+P HT use given that E+P HT use has been suggested to increase the risk of breast cancer.[31] This potential channeling bias would have led to higher rates of breast cancer in the CE/BZA group. What was observed in this group, however, were lower rates relative to E+P HT (IRR_{pooled}=0.79; 95% CI: 0.58, 1.05; HR_{pooled}=0.78; 95% CI: 0.55, 1.10). This finding is of public health importance given that breast cancer is the most common cancer in women in the US,[32] and until CE/BZA was approved, women were faced with a choice between unopposed estrogen, which could increase the risk of endometrial cancer, and E+P HT, which could increase the risk of breast cancer.[8-10,32]CE/BZA could represent an alternative that does not increase the risk of breast cancer relative to E+P HT, though the impact on endometrial cancer remains relatively uncertain due to few endometrial cancer cases and limited follow-up time across the five study databases. In contrast, breast cancer incidence and mortality are considerably more common than endometrial cancer[33] and as a result, this study was able to estimate the

effect of CE/BZA versus E+P HT on breast cancer incidence with greater precision than was available for endometrial cancer.

11.2. Limitations

Limitations to this study are described below. Limitations to generalizability are addressed in Section 11.4.

11.2.1. Unmeasured covariates

As this study was conducted in administrative claims databases, data on behavioral and environmental cancer risk factors were not available, including information on physical activity level, dietary habits, alcohol, tobacco, and environmental exposures. However, at this time, there is no reason to believe that CE/BZA users systematically differ from E+P HT users in their health-related habits and environmental exposures.

Also, while unable to directly measure these factors, this study was able to account for region of residence in all but the smallest database (MarketScan Medicaid). Environmental exposures, health behaviors, and rates of cancer are known to differ by US region.[34,35] Region of residence was nearly identical between CE/BZA and E+P HT users after PS-matching (**Appendices 8a-c & 8e, Table 3**). Controlling for region as done in the analyses may have attenuated confounding by region-related health behaviors such as dietary customs and smoking and alcohol intake, though region is likely only a crude proxy for these individual-level behaviors.

11.2.2. Residual covariate imbalances

Among the confounders that were measured and included, PS-matching was employed to arrive at similar levels of each covariate in each treatment group. After PS matching, the distributions of covariates of interest were largely balanced between the CE/BZA and E+P HT groups in each database and within each cohort. However, there were variables with absolute standardized differences (ASDs) above the conventional threshold of 0.1[36] in multiple databases after matching, particularly in the smaller databases (Healthagen and MarketScan Medicaid). This is of most importance for Cohort 1 (the new user analysis) which served as the basis for the comparative analyses.

In the MarketScan Medicaid database, Cohort 1 (new users) ASDs were >0.1 for age group 40-49 years (25.3% vs. 33.1%), endometrial hyperplasia (0% vs. 1.3%), history of breast, ovarian, or endometrial malignancy (1.3% vs. 0.3%), corticosteroids (25.3% vs. 34.1%), azithromycin (32.0% vs. 26.9%), and glucocorticosteroids (25.3% vs. 34.9%). In the ORD, year of cohort entry remained unbalanced after PS matching, and this variable was added to the final outcome model (Cox model) to adjust for this imbalance.

Covariate imbalance was more common in prevalent users (Cohorts 2 and 3) across databases, and PS matching was not possible for Cohort 3 (prior unopposed estrogen) in two of the databases, further underscoring the inappropriateness of using the prevalent users data (Cohorts 2 and 3) for direct comparisons between CE/BZA and E+P HT. Interpretation of the

incidence rates for Cohorts 2 and 3 was already limited by relatively small cohort sizes, and by the potential for prevalent user bias.[17]

11.2.3. Confounding by indication and protopathic bias

Confounding by indication and protopathic bias represent unique types of bias that are relevant to pharmacoepidemiology studies.[37] CE/BZA in the US is not specifically indicated for women in whom "treatment with progestin-containing therapy is not appropriate" as it is in the EU. Given the side effects and breast cancer risk associated with progestin-containing hormone therapy, US CE/BZA users may be more predisposed to (or concerned about) breast cancer or uterine bleeding (which is a symptom of endometrial hyperplasia or endometrial cancer) than US E+P HT users, and may already have undiagnosed endometrial hyperplasia or endometrial cancer which is causing their uterine bleeding. These scenarios could lead to confounding by indication and protopathic bias (reverse causality). However, after PS-based matching, breast pain or lump, pre-existing endometrial hyperplasia, family history of cancer, and history of breast, ovarian, or endometrial malignancy were balanced between treatment groups in each database for the comparative analyses. With the exception of the Healthagen database, baseline mammography was also balanced (i.e., ASD < 0.1). However, claims-based ascertainment of cancer history – particularly family history of cancer, is likely to be incomplete.

11.2.4. Limited follow-up time

Further limiting this study's ability to detect effects was the limited follow-up time inherent to US claims databases combined with loss-to-follow-up related to Medicare enrollment in the ≥64 years age group (Section 10.5.2). These two factors could result in missing outcomes that take years to develop or to detect, such as cancer outcomes with long induction and latency periods. This is of relevance to studies of endometrial cancer and drug exposure. For example, estrogen modulator tamoxifen has been associated with endometrial carcinoma with a lag of 0.7 to 14 years, with a mean of 24 months reported in one study.[38,39] This study included 44,414 new users of CE/BZA or E+P HT (Cohort 1) with 82,458 person-years of follow-up for the endometrial cancer analysis, or 1.86 years (i.e., less than 24 months) of follow-up per person (average), and captured 39 cases of endometrial cancer.

11.2.5. Outcome validation

While a previously validated algorithm in the HIRD was used to accurately identify endometrial hyperplasia cases (PPV=80%) and endometrial cancer (PPV=91%),[23,24] the transportability of these algorithms to the other four databases is uncertain because other databases could have different underlying prevalences of endometrial hyperplasia and cancer (note: PPV is affected by prevalence). The quantitative bias analysis described in Section 10.5.1.3 suggests that differential outcome misclassification could have impact on the study estimates, although there is no current reason to believe that misclassification would be large enough to change the conclusions from this study, as described in Section 10.5.1.3. The algorithms for the secondary outcomes were not validated specifically in any of the five utilized databases, and the proportion of false positives and negatives across databases is unknown.

11.2.6. Population heterogeneity

Pooling of data from the HIRD, Healthagen, and ORD is intuitive given that all three are claims databases for major US insurance companies and that these data represent patients with employer-sponsored insurance. However, each database, as described in Table 3, is unique in its dates of data availability, and inclusion or exclusion of administrative services only (ASO) and Medicare advantage patients. Furthermore, each database represents a different collection of employers, and differences in occupational exposure may be relevant to the outcomes under study. As a result, there may be differences in the patient population that are pertinent to this study that are unaccounted for and unmeasurable by claims data. The MarketScan Medicaid population might be thought of as a sample from a different population altogether, given the higher morbidity and poverty inherent to this population and related unmeasurable factors. As a result, pooled findings represent a qualitatively heterogeneous mixture of patients in terms of clinical comorbidities, age, and other factors. However, each database can be thought of as a component of the total population taking CE/BZA or E+P HT. Further, given the large size of these databases representing over 90 million women – study findings represent a large proportion of women in the US receiving these medications.

11.2.7. Comparator

The E+P HT comparator was composed of a variety of estrogen and progestin molecules. The type of estrogen and type of progestin could produce different side effect profiles and potentially different risks for the outcomes under study. The type of progestin which CE/BZA is compared to, may be of particular importance to driving comparative estimates, particularly for breast cancer.[40] In other words, this study may have reported different findings had the composition of the comparator group in terms of type of E+P HT been different. The bulk of the E+P HT group were users of CE/medroxyprogesterone (n=20,500; 27%) or estradiol/norethindrone acetate (n=18,486; 24%). Norethindrone acetate, a first-generation progestin with moderate androgenic activity, was the most common progestin type (n=32,434; 43%) and estradiol the most common estrogen type (n=31,240; 41%) in the E+P HT users (Table 3).

11.3. Interpretation

The inclusion of progestin in E+P HT is intended to protect patients against endometrial cancer,[7] yet the inclusion of progestin may confer additional risk for breast cancer.[8-10] Until CE/BZA was approved, women were faced with a choice between unopposed estrogen, which could increase the risk of endometrial cancer, and E+P HT which could increase the risk of breast cancer.[31] CE/BZA may represent an alternative that does not increase the risk of breast cancer compared to E+P HT, though the impact on endometrial cancer remains relatively uncertain due to few endometrial cancer cases and limited follow-up time across the five study databases. In contrast, breast cancer is more common than endometrial cancer and as a result, the study was able to estimate the effect of CE/BZA vs. E+P HT on breast cancer incidence with greater precision than was available for endometrial cancer.

Importantly, overall, the risks of cancer to any site and the risks of acute cardiovascular outcomes were similar in CE/BZA new users and E+P HT new users (**Appendix 8f; Table 11**), potentially suggesting a relatively similar overall safety event profile at the population level.

11.4. Generalizability

11.4.1. Generalizability to EU population

Oral conjugated estrogens/bazedoxifene (CE/BZA) was authorized in the US for the treatment of moderate to severe vasomotor symptoms associated with menopause and prevention of postmenopausal osteoporosis in women with a uterus.[41] In the EU, CE/BZA is not indicated for the prevention of osteoporosis and is specifically indicated for women in whom "treatment with progestin-containing therapy is not appropriate."[42] As a result, the indicated population for CE/BZA differs between the US and the EU. In the EU, there may be fewer women who take CE/BZA who are concerned with osteoporosis risk, for example.

The incidence rate ratios and hazard ratios produced by this study should still be generalizable to the EU CE/BZA users as long as the differences in the EU indications are not modifiers of the effect of CE/BZA on any of the study outcomes. At this time, there is no reason to believe that using CE/BZA for osteoporosis prevention as opposed to vasomotor symptoms modifies the effect of CE/BZA on any of the outcomes that were studied, particularly in comparison with E+P HT.

Less apparent, however, are the implications of the EU's stipulation that the drug is indicated specifically for women in whom *"treatment with progestin-containing therapy is not appropriate."* It is less clear how this is interpreted by providers and patients, though it seems to imply women who are at higher breast cancer risk. Hormone therapy formulations often include progestin (synthetic progesterone) to protect patients from endometrial cancer.[7] However, progestin may be avoided by women at higher breast cancer risk[8-10] or who experience side effects from progestin, such as abnormal uterine bleeding. Due to these concerns, there was a market need for an alternative approach to estrogen opposition in hormone therapy, and CE/BZA may be used as an alternative to E+P HT in these women. As a result, CE/BZA users may be more predisposed to (or concerned about) breast cancer or uterine bleeding (which is a symptom of endometrial hyperplasia or cancer) or may already have undiagnosed endometrial hyperplasia or cancer which is causing their uterine bleeding. In the US, women do not need a specific reason to use CE/BZA instead of E+P HT, and therefore the population may be broader than the more narrow population of CE/BZA users in the EU who may be more likely to develop these outcomes, and in whom the contrast between risks in CE/BZA and E+P HT may be more or less pronounced.

11.4.2. Other generalizability stipulations

The study findings were estimated in a mostly commercially insured (employer-based) population, with a small subset of Medicaid patients (1% of pooled comparative results). Given that the Medicaid database represents over 10 million women during the study period, the small subset of CE/BZA users was likely reflective of the lack of the uptake of the drug

in this population. Additionally, uninsured women, and women exclusively covered by traditional fee-for-service Medicare were not included. While women in these group may be unlikely to initiate CE/BZA due to healthcare access or age-related factors, they may represent different underlying comorbidity and concomitant medication distributions which may modify the effect of CE/BZA vs. E+P HT on the outcomes under study, and thus findings are most relevant to the employer-sponsored insured population and should be interpreted as such.

12. OTHER INFORMATION

Not Applicable.

13. CONCLUSIONS

In this study of five large US claims databases representing over 92 million insured women, CE/BZA use was uncommon, with only 28,837 (0.03%) CE/BZA users identified. The ratios of incidence rates for endometrial cancer and endometrial hyperplasia comparing new CE/BZA users to new E+P HT users were both greater than 1.00 but with a 95% CI that included 1.00, though, this finding was limited by few cases and limited follow-up time across the five study databases. These outcomes were uncommon, thus the study had limited statistical power to estimate small or moderate magnitudes of effect. In contrast, breast cancer is more common than endometrial cancer and as a result, this study was able to estimate the effect of CE/BZA versus E+P HT on breast cancer with greater precision than was available for other outcomes, and demonstrated no increased rate of breast cancer in CE/BZA users relative to E+P HT. Importantly, the risks of cancer to any site and the risks of acute cardiovascular outcomes were similar in CE/BZA new users and E+P HT new users, suggesting a similar risk profile at the population level.

14. REFERENCES

1. Gold EB. The timing of the age at which natural menopause occurs. *Obstet Gynecol Clin North Am.* 2011;38(3):425-440.
2. The American College of Obstetricians and Gynecologists. Frequently Asked Questions: What are Menopause and Perimenopause. 2020; https://www.acog.org/womens-health/faqs/perimenopausal-bleeding-and-bleeding-after-menopause. Accessed 26 October, 2020.
3. The American College of Obstetricians and Gynecologists. Frequently Asked Questions: The Menopause Years. 2020; https://www.acog.org/womens-health/faqs/the-menopause-years. Accessed 27 October, 2020.
4. Avis NE, Crawford SL, Green R. Vasomotor Symptoms Across the Menopause Transition: Differences Among Women. *Obstet Gynecol Clin North Am.* 2018;45(4):629-640.
5. Stoppler MC. Medical Definition of Replacement Therapy, Hormone (HRT, HT). https://www.medicinenet.com/replacement_therapy_hormone_hrt_ht/definition.htm. Accessed 9 March, 2021.
6. Sturdee D, MacLennan A. HT or HRT, that is the question? *Climacteric.* 2003;6(1):1.
7. Beral V, Bull D, Reeves G. Endometrial cancer and hormone-replacement therapy in the Million Women Study. *Lancet.* 2005;365(9470):1543-1551.
8. National Cancer Institute. Menopausal Hormone Therapy and Cancer. 2018; https://www.cancer.gov/about-cancer/causes-prevention/risk/hormones/mht-fact-sheet. Accessed 27 October, 2020.
9. Chlebowski RT, Anderson G, Pettinger M, et al. Estrogen plus progestin and breast cancer detection by means of mammography and breast biopsy. *Arch Intern Med.* 2008;168(4):370-377; quiz 345.
10. Chlebowski RT, Kuller LH, Prentice RL, et al. Breast cancer after use of estrogen plus progestin in postmenopausal women. *N Engl J Med.* 2009;360(6):573-587.
11. Kagan R. The Latest Word on Hormone Therapy. 2020; https://www.sutterhealth.org/health/womens-health/the-latest-word-on-hormone-therapy. Accessed 27 October, 2020.
12. U.S. Food & Drug Administration. Drug Approval Package: Duavee (Conjugated Estrogens and Bazedoxifene). 2013; https://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/022247_duavee_toc.cfm. Accessed 27 October, 2020.
13. Duavee [Package Insert]. In. Philadelphia, PA: Wyeth Pharmaceuticals Inc. A subsidiary of Pfizer Inc.; 2013.
14. European Medicines Agency. Duavive. Oestrogens Conjugated/Bazedoxifene. 2015; https://www.ema.europa.eu/en/medicines/human/EPAR/duavive. Accessed 28 October, 2020.
15. Biskobing DM. Update on bazedoxifene: a novel selective estrogen receptor modulator. *Clin Interv Aging.* 2007;2(3):299-303.
16. Menopause: Medicines to Help You. In: U.S. Food & Drug Administration; 2017.

17. Danaei G, Tavakkoli M, Hernán MA. Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. *Am J Epidemiol.* 2012;175(4):250-262.
18. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* 2008;19(6):766-779.
19. Prentice RL, Pettinger M, Anderson GL. Statistical issues arising in the Women's Health Initiative. *Biometrics.* 2005;61(4):899-911; discussion 911-841.
20. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software.* 2010;36(3):1-48.
21. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial, Evidence-Based Mental Health. *Evidence-Based Mental Health.* 2019;22(4).
22. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology.* Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
23. Esposito D, Yin R, Russo L, et al. Development and Validation of an Algorithm to Identify Endometrial Hyperplasia in US Administrative Claims Data. Abstract. *Pharmacoepidemiol Drug Saf.* 2016;25(S3):3-679.
24. Esposito DB, Banerjee G, Yin R, et al. Development and Validation of an Algorithm to Identify Endometrial Adenocarcinoma in US Administrative Claims Data. *J Cancer Epidemiol.* 2019:1938952. eCollection.
25. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med.* 2009;28(25):3083-3107.
26. Cannas M, Arpino B. A comparison of machine learning algorithms and covariate balance measures for propensity score matching and weighting. *Biom J.* 2019;61(4):1049-1072.
27. Borenstein M, Higgins J, Rothstein H, Hedges L. I2 is not an absolute measure of heterogeneity in a meta-analysis. https://www.meta-analysis-workshops.com/download/common-mistakes1.pdf. Accessed 11/3, 2020.
28. Lee YH. An overview of meta-analysis for clinicians. *Korean J Intern Med.* 2018;33(2):277-283.
29. Brenner H, Gefeller O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. *Am J Epidemiol.* 1993;138(11):1007-1015.
30. Grady D, Gebretsadik T, Kerlikowske K, Ernster V, Petitti D. Hormone replacement therapy and endometrial cancer risk: a meta-analysis. *Obstet Gynecol.* 1995;85(2):304-313.
31. Manson JE, Chlebowski RT, Stefanick ML, et al. Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. *Jama.* 2013;310(13):1353-1368.
32. Henley SJ, Ward EM, Scott S, et al. Annual report to the nation on the status of cancer, part I: National cancer statistics. *Cancer.* 2020;126(10):2225-2249.
33. World Health Organization (WHO) International Agency for Research on Cancer (IARC). Corpus Uteri Fact Sheet. 2018;

https://gco.iarc.fr/today/data/factsheets/cancers/24-Corpus-uteri-fact-sheet.pdf. Accessed 11/4/2020.

34. Centers for Disease Control and Prevention (CDC) Division of Cancer Prevention and Control. U.S. Cancer Statistics Data Visualizations Tool. *United States Cancer Statistics (USCS)* 2020; https://www.cdc.gov/cancer/uscs/dataviz/index.htm. Accessed 11/4/2020.
35. Centers for Disease Control and Prevention (CDC)/National Center for Chronic Disease Prevention and Health Promotion/Office on Smoking and Health. Tobacco Use by Geographic Region. *Smoking & Tobacco Use* 2019; https://www.cdc.gov/tobacco/disparities/geographic/index.htm. Accessed 11/4/2020.
36. Ali MS, Groenwold RH, Belitser SV, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *J Clin Epidemiol.* 2015;68(2):112-121.
37. European Medicines Agency. ENCePP Guide on Methodological Standards in Pharmacoepidemiology. 2020; http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml. Accessed 29 January, 2021.
38. Dhull AK, Kaushal V, Singh S, Sen R. Tamoxifen-induced endometrial carcinoma after a lag of 14 years. *South Asian J Cancer.* 2013;2(1):6.
39. Daniel Y, Inbar M, Bar-Am A, Peyser MR, Lessing JB. The effects of tamoxifen treatment on the endometrium. *Fertil Steril.* 1996;65(6):1083-1089.
40. McCarty MF. Androgenic progestins amplify the breast cancer risk associated with hormone replacement therapy by boosting IGF-I activity. *Med Hypotheses.* 2001;56(2):213-216.
41. Center for Drug Evaluation and Research. Approval Package for: Application Number: 022247Orig1s000. 2013; https://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/022247Orig1s000Approv.pdf. Accessed 2 November, 2020.
42. Duavive: EPAR -- Product Information. Annex 1 Summary of Product Characteristics. 2015; https://www.ema.europa.eu/en/documents/product-information/duavive-epar-product-information_en.pdf. Accessed 2 November, 2020.

15. LIST OF SOURCE TABLES AND FIGURES

15.1. Tables from the SAP (Appendix 4)

Table 1. Study cohort classifications by baseline medication use

Table 3. Censoring variables for each outcome analysis

Table 7. List of covariates and time frame to assess each covariate

SAP Appendix A: Code Lists

SAP Appendix B: Table Shells

SAP Appendix D: Drug Codes

15.2. Tables from Study Results

15.2.1. Pooled Analysis of All Databases (Appendix 8f)

Table 11. Incidence of study outcomes - pooled estimates including all five databases

Table 12. Incidence of study outcomes - pooled estimates excluding MarketScan Commercial Database

Table 13. Sensitivity analysis to assess the impact of misclassification of Endometrial Cancer (13a) / Hyperplasia (13b) in this cohort study (reformatted original table shell)

Table 14. Proportion of patients lost-to-follow-up after turning 65 in each database

15.2.2. HealthCore Site-Specific Results (Appendix 8b)

Table 8a. Incidence and adjusted hazard ratios of study outcomes comparing CE/BZA users to E+P HT users

Document Approval Record

Document Name:	Duavive_B2311060_Final_Report_fv
Document Title:	B2311060

Signed By:	Date(GMT)	Signing Capacity
Campbell, Ulka	29-Mar-2021 11:36:06	Final Approval
De Bernardi, Barbara	29-Mar-2021 12:53:38	Final Approval