Epidemiology of vasomotor symptoms (VMS) in menopausal women (EpiVaSym): a multi-country, large-scale OHDSI network analytic study

Ronald Herrera[1], Victoria Banks[2], Cecilia Caetano[1], Cecile Janssenswillen[1], Carsten Moeller[1], Siir Su Saydam[1], James Brash[3], James Yang[3], Sarah Seager[3], Christian Reich[3], Renate Schulze-Rath[1]

Affiliations

[1]Bayer AG, Berlin, Germany

[2]Bayer plc, Reading, UK

[3]Real World Solutions, IQVIA, USA

Corresponding author: Ronald Herrera; ron.herrera@bayer.com

Word count: 987/1000 [excluding tables, figures, and references]

Background

Approximately 47 million women enter menopause each year globally, with women now spending around a third of their lives in the postmenopause period.[1] Up to 80% of menopausal women will experience vasomotor symptoms (VMS), including hot flashes, which can be debilitating and affect their daily functioning and quality of life.[2]

Real-world journeys of women with VMS are under-researched, and reported prevalence and incidence data for VMS are outdated and vary widely. This may be explained by study design limitations (many are cross-sectional studies with inconsistent symptom definitions), or demographic, cultural, healthcare system, diagnostic coding, and other variables.[3–7]

Large anonymized electronic datasets, including claims data and electronic health records (EHRs), offer opportunities to generate simultaneous standardized real-world epidemiologic data,[8,9] which can address the variability of evidence available on VMS patient journeys, incidence, and prevalence. However, preliminary research has revealed a number of methodologic issues with using such secondary databases, including:

- diagnostic challenges for clinicians and under-reporting by patients, impacting identification of the true incidence of menopause and related symptoms

- misclassification of diagnosis or outcomes due to lack of specificity in coding by healthcare practitioners
- inconsistent coding of symptoms through the chronic course of menopausal symptoms, impacting prevalence rates
- variation in algorithms used to identify menopause and menopausal symptoms across database studies
- selection bias toward women with access to healthcare and seeking medical treatment for their symptoms.

The EpiVaSym study is a large-scale Observational Health Data Sciences and Informatics (OHDSI) network analytics study designed to characterize the journeys of women experiencing VMS in natural menopause, including those not eligible for hormone therapy (HT) due to contraindication, by using common definitions across multiple data sources and countries.

Aims

- To construct and characterize multiple target cohorts of menopausal women, including those ineligible for HT, using different cohort definitions
- To explore differences in the identification, incidence, and prevalence estimates of VMS in menopausal women based on different target and outcome definitions
- To explore the application of statistical processes in the OHDSI platform to refine epidemiologic estimates of VMS

Methods

Study design and data sources

EpiVaSym is a retrospective, multi-country cohort study using six databases including >312 million people from five countries: three observational, healthcare EHR databases from France, Germany, and the UK; and three administrative claims databases, one from Japan and two from the USA (Table 1). Databases are standardized to the Observational Medical Outcomes Partnership common data model, which enables the use of standardized analytics and tools across the network because the structure of the data and the terminology system are harmonized.

Two target study cohorts will include (A) women with a first record of natural menopause (index event) recorded between the ages of 40 and 65 years from January 1, 2010 until December 31, 2019; and (B) women experiencing natural menopause not eligible for HT due to contraindication.

We will use demographic, comorbidity, and concomitant medication covariates to characterize the study target cohorts at multiple time periods in the patient journey relative to the index event (Table 2).

Data analysis

This is a descriptive study without formal hypothesis testing. Counts will be calculated for the number of patients in each database that match the target cohort criteria. Study results will be aggregated statistics, and no individual patients will be identifiable.

Due to the limitations noted above regarding the use of secondary databases for identifying menopause and menopausal symptoms, multiple lists of concepts will be developed to define the target cohort conditions, outcomes, and covariates. These will be built using a combination of published literature and specific READ, SNOMED, and ICD-9/10 codes selected by a multidisciplinary team including clinicians and epidemiologists.

Incidence and prevalence rates will be calculated using the OHDSI ATLAS tool "Incidence Rates". Analyses will be performed using the R package "VMSChar" developed by Bayer using the OHDSI Cohort Diagnostics library.[10] Baseline covariates will be extracted using an optimized SQL extraction script based on the principles of the "FeatureExtraction" package[11] to quantify demographic, comorbidity, and medication groups.

Subject baseline disease characteristics at index date will be reported using medians for non-normally distributed continuous variables, and proportions for categorical variables. Kaplan–Meier analyses will be used to assess time from index date to first coded VMS diagnosis in each target cohort. Women will be stratified according to predefined subgroups (demographics, comorbidities, medications, and specific diagnoses) including those with and without VMS.

Results

A preliminary assessment suggests between 81,657 and 2,413,809 women match the criteria of "natural menopausal women" in the databases used. Incidence and prevalence of VMS estimates are expected to differ depending on the cohort definitions used to characterize women undergoing natural menopause (Table 3).

Conclusions

We have described the methodology for a large simultaneous multi-country cohort study of VMS in menopausal women, including those not eligible for HT, using large-scale network analytics to define multiple target and outcome cohorts.

Anticipated study design strengths include:

- the use of up-to-date real-world data reflecting routine clinical menopause care
- development of multiple target cohorts based on a range of definitions used in the absence of standard definitions of menopause and VMS
- simultaneous standardized and directly comparable multi-country cohorts with potential to observe differences across countries in real-world coding patterns and clinical practice specifically for menopause.[8,9,12–15]

Anticipated study limitations include:

- selection bias in retrospective EHR databases, in which data reflect care received only while enrolled in the healthcare system
- under-reporting of symptoms by patients
- recording and coding reflecting physician perception, motivation, and clinical practice
- challenges in the accurate assessment of the stage and duration of menopause.

The EpiVaSym study will provide insights into the burden of VMS, and the journeys of women with VMS in the healthcare system. It will assist in identifying and characterizing menopausal women using real-world secondary data sources, and establish the basis for consistent coding and algorithm definitions to produce robust epidemiologic estimates. Finally, it will inform the development of further research hypotheses in different patient groups and countries to improve the health of menopausal women.

References

1. Hill K. The demography of menopause. *Maturitas*. 1996 Mar;23(2):113-27.
2. Nappi RE, Kroll R, Siddiqui E, *et al*. Global cross-sectional survey of women with vasomotor symptoms associated with menopause: prevalence and quality of life burden. *Menopause.* 2021 May;28(8):875-82.
3. Li J, Luo M, Tang R, *et al*. Vasomotor symptoms in aging Chinese women: findings from a prospective cohort study. *Climacteric.* 2020 Feb;23(1):46-52.
4. Islam RM, Bell RJ, Rizvi F, Davis SR. Vasomotor symptoms in women in Asia appear comparable with women in Western countries: a systematic review. *Menopause.* 2017 Nov;24(11):1313-22.
5. Richters JM. Menopause in different cultures. *J Psychosom Obstet Gynaecol*. 1997 Jun;18(2):73-80.
6. Freeman EW, Sherif K. Prevalence of hot flushes and night sweats around the world: a systematic review. *Climacteric.* 2007 Jun;10(3):197-214.
7. Gold EB, Sternfeld B, Kelsey JL, *et al*. Relation of demographic and lifestyle factors to symptoms in a multi-racial/ethnic population of women 40-55 years of age. *Am J Epidemiol.* 2000 Sep;152(5):463-73.
8. Hripcsak G, Duke JD, Shah NH, *et al*. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform.* 2015;216:574-8.
9. Schuemie MJ, Ryan PB, Hripcsak G, Madigan D, Suchard MA. Improving reproducibility by using high-throughput observational studies with empirical calibration. *Philos Trans A Math Phys Eng Sci.* 2018 Sep;376(2128):20170356.
10. Rao G, Schuemie M, Ryan P, Weaver J, Gilbert J. *CohortDiagnostics: Diagnostics for OHDSI Cohorts* [Internet]. 2022. Available from: https://ohdsi.github.io/CohortDiagnostics [Accessed on: May 12, 2022].
11. Schuemie MJ, Ryan PB, Suchard MA, Reps J, Sena A. *FeatureExtraction.* [Internet]. 2022. Available from: https://ohdsi.github.io/FeatureExtraction/ [Accessed on: May 12, 2022].
12. Rathmann W, Bongaerts B, Carius HJ, Kruppert S, Kostev K. Basic characteristics and representativeness of the German Disease Analyzer database. *Int J Clin Pharmacol Ther.* 2018 Oct;56(10):459-66.
13. Jouaville SL, Miotti H, Coffin G, Sarfati B, Meihoc A. Validity and limitations of the Longitudinal Patient Database France for use in pharmacoepidemiological and pharmacoeconomics studies. *Value Health.* 2015 May;18(3):A18.

14. Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol.* 2015 Jun;44(3):827-36.

15. Boggon R, van Staa TP, Chapman M, Gallagher AM, Hammad TA, Richards MA. Cancer recording and mortality in the General Practice Research Database and linked cancer registries. *Pharmacoepidemiol Drug Saf.* 2013 Feb;22(2):168-75.

Tables

Table 1. Data sources.

Data source	Source population	Sample size, millions	Providers	Data type	Year first established	Percentage population covered
Disease Analyzer (DA) Germany	Ambulatory	38.5	2734 GPs and specialists	Electronic health records	1992	46%
Longitudinal Patient Database (LPD) France	Ambulatory	17.9	>1200 GPs from 400 practices	Electronic health records	1994	27%
Clinical Practice Research Datalink (CPRD GOLD) UK	Ambulatory	16.25	GPs	Electronic health records	1984	8.5%
OPTUM Clinformatics USA	Closed claims	69.39	NA	Claims data	2007	19% of the insured US population[†]
IBM MarketScan USA	Closed claims	164.78	NA	Claims data	2002	167 million person-years of data
Japan Claims Japan	Closed claims	3.5	NA	Claims data	2013	2.8%

[†]This means 19% of those in commercial health plans and 19% of those in Medicare Advantage plans.

GP, general practitioner; NA, not applicable.

Table 2. Covariates and time restrictions.

Covariate	Time restriction
Demographics	
Age as a continuous variable, years	At index date
Age in 5-year categories, years	At index date
BMI <20 kg/m^2	At index date
BMI 20–24 kg/m^2	At index date
BMI 25–30 kg/m^2	At index date
BMI >30 kg/m^2	At index date
Charlson Comorbidity Index score 0	At index date
Charlson Comorbidity Index score 1	At index date
Charlson Comorbidity Index score ≥2	At index date
Comorbidities	
Alopecia	No restraint
Anxiety	No restraint
Breast cancer	No restraint
Depression	No restraint
Depression without bipolar disorder	No restraint
Gynecologic cancer (e.g. cervical cancer)	No restraint
Hypertension	No restraint
Malaise or fatigue	No restraint
Myocardial infarction	No restraint
Osteoarthritis	No restraint
Osteoporosis	No restraint
Type 2 diabetes	No restraint
Stroke (ischemic or hemorrhagic)	No restraint
Venous thromboembolism	No restraint
Medications	
Antidepressants	No restraint
Aromatase inhibitors	No restraint
GnRH analogs	No restraint
Tamoxifen	No restraint
Tamoxifen or GnRH analogs or aromatase inhibitors	No restraint

BMI, body mass index; GnRH, gonadotropin-releasing hormone.

Table 3. Preliminary assessment of cohort size based on ICD-9/10 codes.

	DA Germany	LPD France	CPRD GOLD UK	IBM MarketScan USA	OPTUM Clinformatics USA	Japan Claims Japan	Total, n
Cohort 1, n[†]	213,202	81,657	146,243	2,413,809	990,665	95,297	3,940,873
Total in database, n	39,194,578	17,864,641	16,225,000	166,710,000	69,389,000	3,481,362	312,864,581

[†]Women with a first record of natural menopause recorded between the ages of 40 and 65 years and on or after January 1, 2010 until December 31, 2019.

CPRD, Clinical Practice Research Datalink; DA, Disease Analyzer; ICD-9/10, International Classification of Diseases 9th/10th Revision; LPD, Longitudinal Patient Database.