

Maternal use of thyroid hormone replacement therapy before, during, and after pregnancy: agreement between self-report and prescription records and group-based trajectory modeling of prescription patterns

Anna S Frank,^{1,2} Angela Lupattelli,¹ David S Matteson,^{2,3} Hedvig Nordeng^{1,4}

¹Pharmacoepidemiology and Drug Safety Research Group, School of Pharmacy, University of Oslo, 0316 Oslo, Norway;

²Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA;

³Department of Statistical Science, Cornell University, Ithaca, NY 14853, USA; ⁴Department of Child Health and Development, National Institute of Public Health, 0403 Oslo, Norway

→ Video abstract



Point your Smartphone at the code above. If you have a QR code reader the video abstract will appear. Or use: <http://youtu.be/KVOhm0nh0Y>

Correspondence: Anna S Frank
Pharmacoepidemiology and Drug Safety
Research Group, School of Pharmacy,
University of Oslo, PO Box 1068 Blindern,
0316 Oslo, Norway
Tel +47 22 85 65 85
Fax +47 2 285 4402
Email a.s.j.frank@farmasi.uio.no

Purpose: A reliable definition of exposure and knowledge about long-term medication patterns is important for drug safety studies during pregnancy. Few studies have investigated these measures for thyroid hormone replacement therapy (THRT). The purpose of this study was to 1) calculate the agreement between self-report and dispensed prescriptions of THRT and 2) classify women with similar adherence patterns to THRT into disjoint longitudinal trajectories.

Methods: Our analysis used data from the Norwegian Mother and Child Cohort Study (MoBa), a prospective population-based cohort study. MoBa was linked to prescription records from the Norwegian Prescription Database (NorPD). We estimated Cohen's kappa coefficients (κ) and approximate 95% CIs for agreement between self-report and prescription records for the 6-month period prior to pregnancy and for each pregnancy trimester. Using group-based trajectory models (GBTMs), we estimated adherence trajectories among women who self-reported and had a THRT prescription.

Results: There were 56,148 women in MoBa, who had both a record in NorPD and available prescription history up to 1 year prior to pregnancy. Of these, 1,171 (2.1%) self-reported and received a prescription for THRT. Agreement was "perfect" in the 6-month period prior to pregnancy ($\kappa=0.86$; CI 0.85–0.88), in the first ($\kappa=0.83$; CI 0.82–0.85) and in the second trimesters ($\kappa=0.89$; CI 0.87–0.90), while this was moderate ($\kappa=0.57$; CI 0.54–0.59) in the third trimester. Among the subset of the 1,171 women, we identified four disjoint GBTM adherence groups: Constant-High (50.2%), Constant-Medium (32.9%), Increasing-Medium (11.0%), and Decreasing-Low (5.8%).

Conclusion: Agreement between self-report and prescription records was high for THRT in the early pregnancy period. Based on our GBTM results, about one in two women with hypothyroidism had adequate adherence to prescribed THRT throughout pregnancy. Given the potential consequences, evidence of low adherence in 5.8% of pregnant women with hypothyroidism is of concern.

Keywords: group-based trajectory models, κ , hypothyroidism, pregnancy, MoBa, NorPD

Introduction

Pregnancy encompasses many physiological changes that profoundly impact thyroid function and creates an increased need for the production of thyroid hormones.¹ During an uncomplicated pregnancy, thyroid hormone levels adapt automatically;

however, women with hypothyroidism require adequate and continuous dosage of THRT throughout pregnancy.¹ The benefits of adequate THRT for maternal and child health have been reported in several studies.¹ For example, Abalovich et al² analyzed the effect of inadequate THRT dosage on prematurity among 150 pregnant Argentinian women.² Of the 123 women with adequate treatment at conception and during pregnancy, 1.6% had a premature birth, compared to 12.5% of the 27 inadequately treated women.²

When data on thyroid hormone blood levels are insufficient or outright unavailable, researchers have relied on alternative data to quantify the effect of THRT on pregnancy outcomes.³ As a result, the need for data reliability assessment arises, particularly in relation to THRT exposure.⁴ The accuracy of data can be determined when exposure information is derived from several data sources.⁴ A measure of agreement among data sources commonly used in epidemiology is the Cohen's kappa coefficient (κ).⁵ Palmsten et al,⁶ for example, compared the agreement between medical records and maternal self-report for medications for rheumatoid arthritis and asthma using κ .⁶ In pharmacoepidemiology, agreement quantification is important to assess the risk of exposure misclassification. A reliable source of exposure information can reduce bias of effect estimates in safety studies. Despite the importance of THRT for maternal-child health, no previous study has explored the agreement of THRT among Norwegian data sources, and there is limited knowledge about long-term prescription patterns to THRT among pregnant women.

The PDC by a dispensed drug therapy informs how well a patient adhered to a treatment.⁷ Though widely used in pharmacoepidemiology, a disadvantage of PDC is that adherence is measured as a single scalar and does not capture time-varying patterns of medication use. A useful alternative to PDC is GBTM.^{8–10} Based on finite-mixture models, GBTM splits the distribution of longitudinal data into a finite number of disjoint trajectory groups. Though GBTMs are common in psychology or criminology, their use in pharmacoepidemiological studies, and specifically pregnancy studies, is rather rare.

The aim of the current study is therefore to 1) assess the agreement between self-report and prescription records of THRT prior to and during pregnancy, and to 2) use a GBTM approach to cluster women with similar prescription patterns of THRT into longitudinal trajectory groups from 6 months prior to 12 months after pregnancy.

Materials and methods

The present analysis is based on three major data sources: MoBa, MBRN, and NorPD.

MoBa is a prospective, population-based cohort study of pregnancies in Norway that was initiated in 1999 by the Norwegian Institute of Public Health; follow-up is ongoing.¹¹ From 1999 to 2008, all women in Norway were invited to participate through a postal invitation in relation to the routine ultrasound examination around gestational week 17. The participation involved response to MoBa Q1. Of the invited women, around 41% consented to participate, with a 95% response rate in MoBa Q1 at gestational week 17, 92% response rate in MoBa Q3 at gestational week 30, and 87% response rate in MoBa Q4 at 6 months after delivery.¹² The cohort now includes 114,500 children along with 95,200 mothers and 75,200 fathers.¹³ The current study is based on Version 10 of the quality-assured data that were released for research purposes in 2017.

MBRN is a nationwide health registry of information about all births in Norway.¹⁴ MBRN registers all pregnancies ending after gestational week 12, including terminations.¹⁴ The registry includes confirmed medical records related to maternal health before and during pregnancy, including perinatal complications.¹⁴

NorPD is a nationwide prescription registry established in January 2004. Since then, all pharmacies in Norway have been obliged to send data electronically to the Norwegian Institute of Public Health on all prescribed drugs dispensed to individuals in ambulatory care.¹⁵ MoBa, MBRN, and NorPD were combined using the unique pregnancy identification number. The timeline of the MoBa questionnaires and prescription records (NorPD) is illustrated in Figure S1.

Definitions of THRT and maternal characteristics

Information on maternal self-report of THRT was ascertained from the questionnaire-specific item “thyroid disorder” in MoBa Q1.¹⁶ MoBa Q1 covers the 6-month period prior to pregnancy and the first 4 months during pregnancy (ie, gestational weeks 0–4, 5–8, 9–12, and 13+). In MoBa Q3 and MoBa Q4, women could report THRT use in weeks 13–32, as well as in the last part of pregnancy, under the “other medications” section. THRT was classified based on the ATC Classification System and included levothyroxine (ATC code H03AA01) and liothyronine (ATC code H03AA02).¹⁷

We created binary medication exposure for each questionnaire interval (6 months prior to gestation, weeks 0–4

to 25–28, interval 29+, and last part of pregnancy), as self-reported by women checking the questionnaire boxes.

For first-trimester exposure, we combined gestational weeks 0–4, 5–8, and 9–12. Use of THRT in at least one of these intervals classified the respondent as exposed during the first trimester, and as unexposed, if none of them was marked. The second trimester combined gestational weeks 13–16, 17–20, 21–24, and 25–28, and exposure was defined similar to the first trimester. The third trimester included the intervals 29+ and last part of gestation from, respectively, MoBa Q3 and MoBa Q4. Exposure in the third trimester, and during the whole gestational period, was defined similar to the other two trimesters.

NorPD provided information about THRT-dispensing dates and the amount of DDDs. The date of last menstrual period and length of gestation were ascertained from MBRN. Based on these, we defined the following time periods: 6-month pregestational period, 4-week gestational intervals, trimester, pregnancy period, and 12-month postnatal period. Next, based on dispensing date and the number of DDDs, we calculated the supply of THRT.

We then defined a binary medication exposure variable for THRT (ATC code H03AA) for each time period: a woman was classified as exposed in a time period if dates of supply overlapped, by at least 1 day, within the corresponding time period. After we defined start and end dates for each dispensed prescription, we also calculated medication gaps during the pregnancy period. A medication gap was present if a woman did not receive a prescription for THRT for at least 14 consecutive days during pregnancy.¹⁸ In addition to medication gaps, we determined the precise amount of DDDs dispensed to each woman.

Information about sociodemographic characteristics was obtained from MoBa Q1, including education, income, BMI at conception, and pregnancy planning, and from MBRN, including maternal age, marital status, smoking habits in early pregnancy, and parity. The lifetime history of major depression (LTHMD) was measured in the MoBa Q1 by Kendler et al's lifetime major depression scale, including five items that closely correspond to the Diagnostic and Statistical Manual of Mental Disorders, Third Edition, criteria for LTHMD.¹⁹ Reproductive history was self-reported in MoBa Q1 and included previous pregnancy outcomes. Items about the perinatal use of recommended nutritional supplements included vitamin D, folic acid, and/or omega-3 fatty acids, either alone or in combination with additional supplements, and were ascertained from MoBa Q1.

Comorbidity variables were classified as medicated or non-medicated, depending on whether the woman reported

treatment for epilepsy (ATC code N03A), arthritis (ATC code M01, L04A, N02), diabetes types I and II (A10A, A10B, A10X), anemia (B03A, B03B, B03X), or cardiovascular disorders (C01–C10) on the MoBa Q1. Mental comorbidity (depression and/or anxiety) was determined by the MoBa Q1 and was categorized as medicated or non-medicated, depending on whether the woman reported psychotropic drug use (ATC codes N05 and N06).

Study population

The study population included women who had a record in MBRN, were enrolled in MoBa Q1, filled MoBa Q3 before delivery, were successfully linked to NorPD, had reported the use of THRT, and received a dispensed prescription (Figure 1). In order to ensure that we captured all prescriptions filled to mothers, we excluded women entering MoBa before the establishment of NorPD in 2004. To capture prescriptions dispensed 6 months prior to pregnancy, we restricted the study population to women filling MoBa Q1 and MoBa Q3 in 2005 and afterward. The potential hypothyroid population consisted of women with THRT exposure record in MoBa or NorPD. The hypothyroid population comprised women with THRT exposure recorded in both data sources. The latter group of women had a record of THRT dispensed from pharmacies in NorPD and self-reported use of THRT in MoBa. The flowchart to achieve the final population is depicted in Figure 1.

Statistical analysis

Descriptive statistics were calculated. We calculated κ and approximate 95% CIs to estimate the agreement between self-report and prescription records.⁵

Agreement was determined in the potential hypothyroid population ($n=1,388$). Agreement was classified according to Landis and Koch as 1) slight, for $\kappa \leq 0.20$; 2) fair, for $0.21 \leq \kappa < 0.40$; 3) moderate, for $0.41 \leq \kappa < 0.60$; 4) substantial, for $0.61 \leq \kappa < 0.80$; and 5) perfect, for $\kappa \geq 0.80$.⁵ κ were calculated for the entire pregnancy period, for the 6-month pregestational period, each trimesters, and 4-week intervals (0–4 to 25–28 weeks). Agreement by maternal factors, such as smoking, education, and BMI, related to the 6 months prior through the end of pregnancy period. Prescription records were considered as the reference standard because of the following: 1) women were specifically asked about “thyroid disease” only in MoBa Q1; 2) MoBa Q1 and MoBa Q3 overlapped for the assessment of exposure around weeks 13–16 but with different questions (“thyroid disease” treatment in MoBa Q1 vs “other medications” in MoBa Q3); and 3) prior

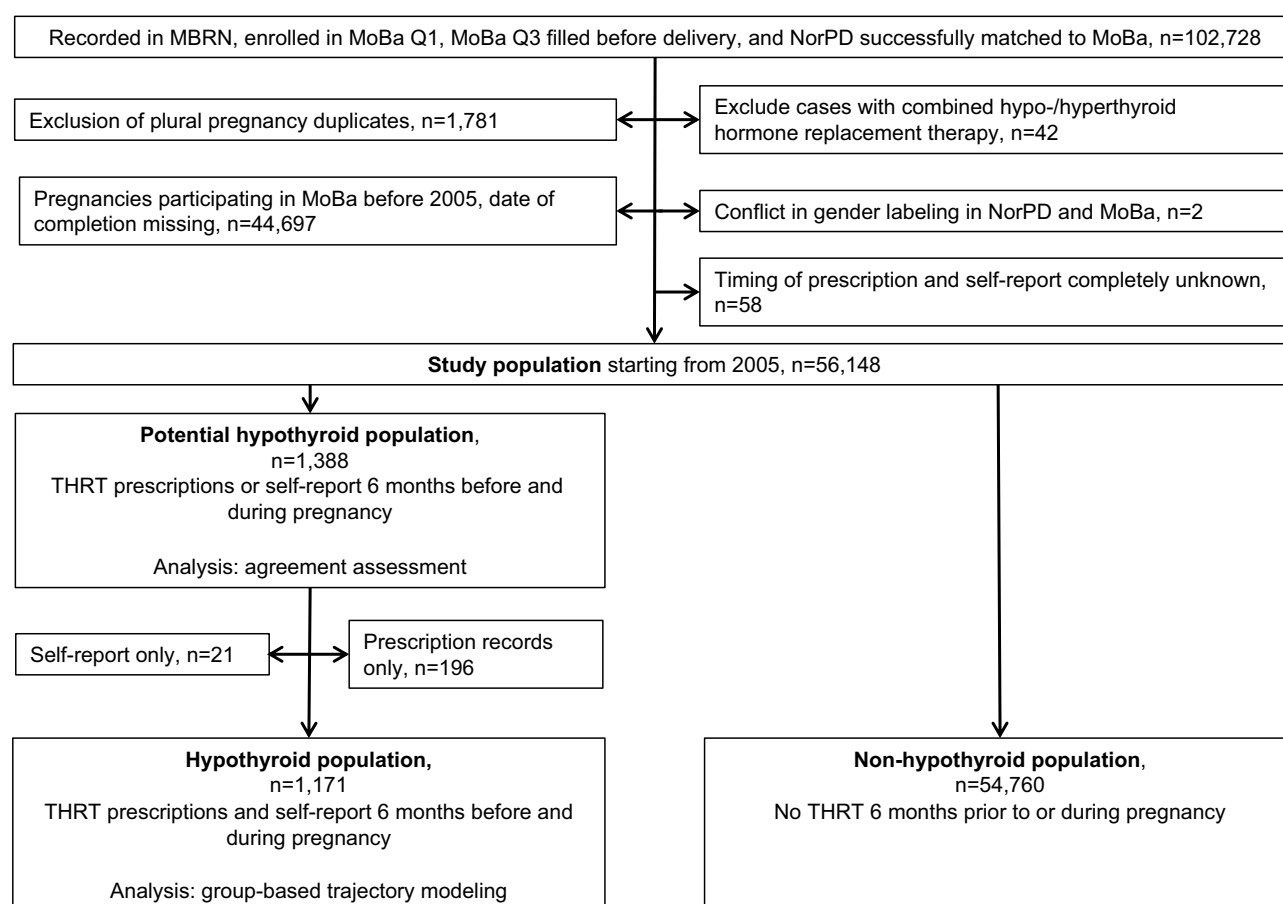


Figure 1 Flowchart of the study population.

Abbreviations: MBRN, Medical Birth Registry of Norway; MoBa, Norwegian Mother and Child Cohort Study; MoBa Q1, MoBa questionnaire 1; MoBa Q3, MoBa questionnaire 3; NorPD, Norwegian Prescription Database; THRT, thyroid hormone replacement therapy.

research used prescription records or medication diaries as reference standard.^{20–23} Values for sensitivity, specificity, PPV and NPV, and 95% CIs were estimated based on Table S1. GBTMs were used to identify women with similar adherence patterns and to cluster them into disjoint trajectory groups.²⁴ For our GBTM analysis, we considered 28 monthly (4-week) binary indicators of dispensed prescriptions. Based on dispensing date and the amount of DDDs dispensed per prescription, we calculated start and end dates of supply. If the dates of supply overlapped with a certain 4-week interval, we coded this as “1” to reflect that prescription records were present for that particular month. If the dates and interval did not overlap, we coded that as “0”.

We calculated supplies of THRT from 6 months prior to pregnancy and up to 12 months after delivery in the hypothyroid population (1,171 women). We restricted the GBTM analysis to the hypothyroid population since only women using THRT at some point around pregnancy can be clustered according to patterns of medication adherence.

In order to identify maternal characteristics specific to the least compliant group of women, we performed multivariate logistic regression for the binary outcome “the least compliant group” vs “the other three groups”, identified by GBTM analysis. There was 27.7% missing information among the covariates. To avoid reduction of the sample size, we carried out multiple imputation for the multivariate analysis, under the assumption that data were missing at random. Further details on multiple imputation procedure are described in the Supplementary materials. Multiple regression analysis with generalized estimating equations was performed on each imputed set. The final-effect estimates were calculated by averaging over effect estimates of all imputed sets.

Data preparation and agreement estimation were performed in R (version 3.4.1). GBTMs were estimated in Stata/MP 15.1 using the “traj” plugin.²⁵ Multiple imputation was performed in R with the package “mice”, and multiple regression analysis was performed using the “survey” package in

R.^{26,27} Model selection on each imputed data set was done using the “step”-function in R.

Ethics

For using data from MoBa, a license was obtained from the Norwegian Data Inspectorate and approval from the Regional Committee for Medical Research Ethics. The overall MoBa study has been approved by the Norwegian Data Inspectorate (01/4325) and the Regional Committee for Medical Research Ethics (S-97045, S-95113). The current study was approved by the Regional Committee for Medical Research Ethics (2015/1241, REK Sør-Øst B). All participants provided written informed consent prior to participation.

Results

The study population comprised 56,148 pregnancy records, which were enrolled in MoBa after 2005 (Figure 1). Around 44.3% of pregnancies were excluded from the study because they were recruited into MoBa before 2005. The characteristics of the study population were similar to those of the overall MoBa population (results not shown). The characteristics of the population by hypothyroid status are outlined in Table 1.

Women with hypothyroidism ($n=1,171$) were older (>30 years, 75.1% vs 57.9%), tended to have a higher BMI at conception (>30 , 17.6% vs 9.7%), and were less often pregnant with the first child (primiparity, 42.4% vs 47.5%) compared to women with no hypothyroidism (Table 1). In addition, women with hypothyroidism complied more often with recommended nutritional supplements (56.2% vs 44.9%), had a higher rate of LTHMD (36.2% vs 23.6%), and were more frequent in the very-high-income class ($>54,443$ USD, 16.9% vs 14.5%).

The rate of women using THRT in both the pregestational and gestational period was approximately 2% (Table 2). The percentage from prescription records was slightly higher than from self-report, and this appeared to be constant over the trimesters. However, during the third trimester, only 1.0% of women self-reported THRT, while 2.1% filled a THRT prescription.

Results of the agreement analysis are presented in Tables 2 and S2. The κ in the whole gestational period was 0.91 (95% CI: 0.89–0.92) (Table 2) and ranged from 0.76 (0.76–0.77) to 0.39 (0.38–0.40) for the weeks 0–4 and 25–28, respectively (Table S2).

The rate of women reporting THRT in MoBa ranged from 1.7% to 1.9% for 4-week intervals 0–4 to 13–16 and dropped to 0.6% for the intervals 17–20 to 25–28 (Table S2). The

percentage of filled prescription records ranged from 1.6% to 1.8% for all 4-week intervals.

Lower κ were associated with being underweight (0.87 [0.77–0.97]), having low educational level (<9 years) (0.75 [0.58–0.92]), as well as smoking at the beginning of pregnancy (results not shown); daily smokers had a κ of 0.91 (0.85–0.96), and sporadic smokers had a κ of 0.84 (0.69–0.98). In comparison, the κ were higher for women being of normal weight (0.92 [0.90–0.93]), overweight (0.92 [0.89–0.94]), or obese (0.91 [0.88–0.94]), for women having medium (9–12 years) (0.88 [0.86–0.91]), high (13–16 years) (0.92 [0.90–0.94]), or very high (>16 years) (0.94 [0.92–0.96]) educational level, and for nonsmokers in early pregnancy (0.92 [0.91–0.93]).

With prescription records as reference standard, sensitivities ranged from 84.5% (82.5%–86.5%) for the whole pregnancy period to 42.3% (39.5%–45.1%) in the third trimester (results not shown). Specificities varied slightly between 99.8% and 99.9%. The PPVs varied from 98.2% (97.4%–98.9%) in the whole gestational period to 88.5% (85.8%–91.1%) in the third trimester, while the NPVs varied from 99.6% (99.5%–99.7%) for the whole pregnancy period to 98.7% (98.7%–98.8%) in the third trimester.

We analyzed adherence trajectory models with one to five groups among the hypothyroid population of 1,171 women. The final four-group model (Figure 2) was chosen as the best fitting model according to the Bayesian Information Criterion under the minimum group size constraint ($>5.0\%$) (Table S3). Four patterns of THRT adherence during the 28-month period around gestation emerged: constant-high (C-H), constant-medium (C-M), increasing-medium (I-M), decreasing-low (D-L). Each woman was assigned to a group based on her maximal membership probability over the groups. The C-H group was the largest, consisting of about 50.2% of all women with hypothyroidism, followed by the C-M group with 33.0% and by the I-M group with 11.0%. The remaining 5.8% were classified as the D-L group.

There were differences in terms of BMI, age, reproductive history, education, income, smoking, and LTHMD between the adherence groups (Table 3).

Variations in aspects of drug utilization between the adherence groups are presented in Table 4: the C-H group had on average more than three prescriptions of THRT with DDDs covering 82% of gestation during pregnancy, while the D-L group had less than two prescriptions of THRT with DDDs covering 20% of gestation. Medication gaps longer than 14 days were most frequent in the C-M group, and the

Table I Characteristics of analytical populations

Variables	Hypothyroid population ^a (n=1,171, %)	Non-hypothyroid population ^b (n=54,760, %)	P-value ^c
Maternal age			
≤24	60 (5.1)	5,622 (10.3)	<0.001
25–29	230 (19.6)	17,385 (31.7)	
30–34	562 (47.9)	21,552 (39.3)	
≥35	319 (27.2)	10,201 (18.6)	
BMI at conception kg/m ²			
≤18	21 (1.8)	1,709 (3.1)	<0.001
19–24	606 (51.7)	33,464 (61.1)	
25–29	308 (26.3)	12,976 (23.7)	
≥30	206 (17.6)	5,316 (9.7)	
Married/cohabitant			
Yes	1,114 (95.1)	52,328 (95.6)	0.517
No	57 (4.8)	2,431 (4.4)	
Pregnancy planning			
Yes	939 (80.2)	44,104 (80.5)	0.699
No	221 (18.9)	10,060 (18.4)	
Parity			
Multiparity	675 (57.6)	28,642 (52.3)	<0.001
Primiparity	496 (42.4)	26,034 (47.5)	
Maternal income ^d			
<16,013 USD	277 (23.6)	12,922 (23.6)	0.048
16,013–54,443 USD	662 (56.5)	32,135 (58.7)	
>54,443 USD	199 (16.9)	7,937 (14.5)	
Educational level-ongoing			
<9 years	12 (1.0)	916 (1.7)	0.080
9–12 years	275 (23.5)	13,770 (25.1)	
12–16 years	487 (41.6)	22,766 (41.6)	
>16 years	377 (32.2)	16,278 (29.7)	
Smoking status in early pregnancy			
Yes ^e	67 (5.7)	3,947 (7.3)	0.058
No	866 (73.9)	39,974 (72.9)	
Use of recommended supplements (MoBa Q1)			
Yes	658 (56.2)	24,579 (44.9)	<0.001
No	513 (43.8)	30,181 (55.1)	
LTHMD			
Yes ^f	424 (36.2)	12,904 (23.6)	<0.001
No	730 (62.3)	40,449 (73.8)	
History of negative reproductive events ^g			
Yes	409 (34.9)	17,789 (32.5)	0.112
No	745 (63.6)	35,759 (65.3)	
Somatic comorbidity (MoBa Q1)			
No	888 (75.8)	48,887 (89.3)	<0.001
Medicated	147 (12.5)	2,147 (3.9)	
Non-medicated	136 (11.6)	3,726 (6.8)	
Mental comorbidity (MoBa Q1)			
No	1,022 (87.3)	49,510 (90.4)	<0.001
Medicated	52 (4.4)	1,152 (2.1)	
Non-medicated	97 (8.3)	4,098 (7.5)	

Notes: ^aThere was missing covariates information in 27.7% of the sample: education (1.7%), pregnancy planning (0.9%), BMI (2.6%), income (2.8%), LTHMD (1.5%), negative reproductive history (1.5%), and smoking (20.3%). ^bThere was missing covariates information in 28.7% of the cases: education (1.9%), pregnancy planning (1.1%), parity (0.1%), BMI (2.4%), income (3.2%), LTHMD (2.6%), negative reproductive history (2.2%), and smoking (19.8 %). ^cChi-square test comparison between non-hypothyroid population and hypothyroid population, rejecting the null hypothesis that the two sets are similar with 0.05 significance level. ^dWomen's income status (USD/year): 1.00 NOK ≈0.13 USD. ^eSmoking consists of daily and occasional smokers. ^fIncludes both major depression without and with negative lifetime event. ^gPrevious negative event includes spontaneous abortion/stillbirth, ectopic pregnancy, and termination of pregnancy (MoBa Q1).

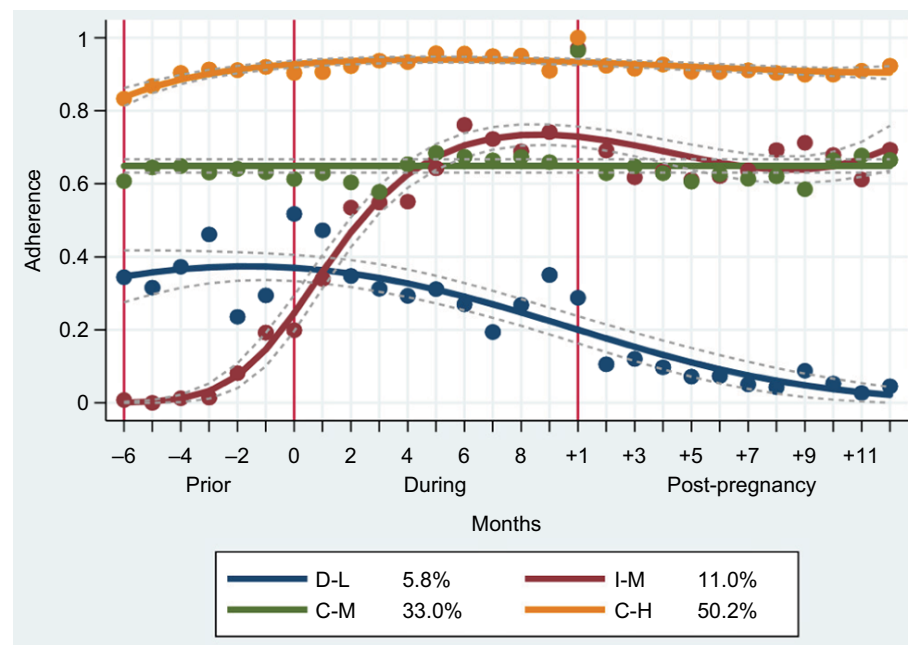
Abbreviations: BMI, body mass index; LTHMD, lifetime history of major depression; MoBa Q1, Norwegian Mother and Child Cohort Study questionnaire 1; NOK, Norwegian Kroner.

Table 2 Agreement between MoBa and NorPD by trimester, pregestational period, and gestational period

THRT	MoBa total ^a		NorPD total ^a		MoBa only	NorPD only	Both	Neither	Agreement	
	n	%	N	%	n	n	n	n	κ^b (95% CI)	Range ^c
6 months prior to LMP	1,001	1.8	1,198	2.1	48	245	953	54,902	0.86 (0.85–0.88)	P
First trimester	1,004	1.8	1,129	2.0	111	236	893	54,908	0.83 (0.82–0.85)	P
Second trimester ^d	1,077	1.9	1,224	2.2	56	203	1,021	54,868	0.89 (0.87–0.90)	P
Third trimester	557	1.0	1,165	2.1	64	672	493	54,919	0.57 (0.54–0.59)	m
Gestational period ^e	1,141	2.0	1,325	2.3	21	205	1,120	54,802	0.91 (0.89–0.92)	P

Notes: ^aTotal study population, n=56,148. ^bCohen's kappa coefficient (κ). ^cSubdivision of agreement by Landis and Koch: <0, no (no); 0–0.20, slight (s); 0.21–0.40, fair (f); 0.41–0.60, moderate (m); 0.61–0.80, substantial (S); and 0.81–1.00, perfect (P). ^dWeeks 13–16 aggregate the text box “week 13+” from MoBa Q1 and “weeks 13–16” from MoBa Q3. ^eLMP to delivery.

Abbreviations: LMP, last menstrual period; MoBa, Norwegian Mother and Child Cohort Study; MoBa Q1, MoBa questionnaire 1; MoBa Q3, MoBa questionnaire 3; NorPD, Norwegian Prescription Database; THRT, thyroid hormone replacement therapy.

**Figure 2** Adherence patterns of four-group trajectories.

Notes: Estimated adherence trajectories (solid lines), observed group means at each month (dot symbols), and estimated percentages for each group: D-L, I-M, C-M, and C-H. Dashed lines are approximated 95% pointwise CIs on the estimated trajectories. Vertical red lines mark the start of the 6-month period prior to pregnancy (at month “-6”), the start of the pregnancy period (at month “0”), and the start of postnatal period (at month “+1”).

Abbreviations: C-H, Constant-High; C-M, Constant-Medium; D-L, Decreasing-Low; I-M, Increasing-Medium.

I-M group increased dosage of THRT in the early months of gestation.

Based on multiple logistic regression analysis, women within the D-L group were four times more likely to have <9 years of education (OR 4.10, 95% CI: 1.05–16.03) compared to the other groups (Table 5).

Discussion

This study is, to our knowledge, the first to estimate the agreement between self-report and prescription records for THRT based on Norwegian data sources. The results show that self-report and prescription records are consistent with especially high agreement in the first and second trimesters.

In the third trimester, agreement between MoBa and NorPD for THRT was only moderate. However, by splitting first and second trimester into 4-week intervals, a clear drop in agreement from MoBa Q1 to MoBa Q3 emerged.

Agreement in the second trimester might be higher than that in the first or third trimester, given that we have defined the former to include one additional 4-week interval. However, the increment in agreement introduced by a sole, additional 4-week interval is considered to be minor and to not substantially change the interpretation of the results. The sensitivity for self-reported data on THRT was very good for the whole gestational period (84.5%), albeit lower for the third trimester (42.3%).

Table 3 Maternal characteristics among the adherence groups

Variables	Adherence groups			
	Decreasing-Low, n=70 (5.8%)	Increasing-Medium, n=128 (11.0%)	Constant-Medium, n=381 (32.9%)	Constant-High, n=592 (50.2%)
Maternal age				
≤24	6 (0.9)	13 (10.1)	27 (7.0)	14 (2.4)
25–29	20 (28.6)	31 (24.2)	81 (21.2)	98 (16.6)
30–34	30 (42.9)	54 (42.2)	179 (46.9)	299 (50.5)
≥35	14 (20.0)	30 (23.4)	94 (24.7)	181 (30.6)
BMI at conception kg/m ²				
≤18	1 (1.4)	4 (3.1)	9 (2.4)	7 (1.2)
19–24	35 (50.0)	66 (51.6)	224 (58.8)	281 (47.5)
25–29	25 (35.7)	33 (25.8)	92 (24.1)	158 (26.7)
≥30	8 (11.4)	22 (17.2)	48 (12.6)	128 (21.6)
Married/cohabitant				
Yes	67 (95.7)	122 (95.3)	362 (95.0)	563 (95.1)
No	3 (4.3)	6 (4.7)	19 (5.0)	29 (4.9)
Pregnancy planning				
Yes	60 (85.7)	97 (75.8)	307 (80.6)	475 (80.2)
No	10 (14.3)	31 (24.2)	69 (18.1)	111 (18.8)
Parity				
Multiparity	38 (54.3)	64 (50.0)	203 (53.3)	370 (62.5)
Primiparity	32 (45.7)	64 (50.0)	178 (46.7)	222 (37.5)
Maternal income ^a				
<16,013 USD	22 (31.4)	30 (23.4)	96 (25.2)	129 (21.8)
16,013–54,443 USD	34 (48.6)	63 (49.2)	218 (57.2)	347 (58.6)
>54,443 USD	11 (15.7)	31 (24.2)	61 (16.0)	96 (16.2)
Educational level-ongoing				
<9 years	3 (4.3)	1 (0.8)	4 (1.0)	4 (0.7)
9–12 years	19 (27.1)	30 (23.9)	88 (23.1)	138 (23.7)
13–16 years	26 (37.1)	47 (36.7)	169 (44.3)	245 (41.4)
>16 years	20 (28.6)	49 (38.3)	113 (29.7)	195 (32.9)
Smoking status in early pregnancy				
Yes ^b	3 (4.2)	9 (7.0)	29 (7.6)	26 (4.4)
No	51 (72.8)	85 (66.4)	276 (72.4)	454 (76.7)
Use of recommended supplements (MoBa Q1)				
Yes	40 (57.2)	57 (44.5)	214 (56.2)	347 (58.6)
No	30 (42.8)	71 (55.5)	167 (43.8)	245 (41.4)
LTHMD				
Yes ^c	11 (15.7)	10 (7.8)	41 (10.8)	49 (8.3)
No	58 (82.8)	117 (91.4)	334 (87.7)	534 (90.2)
History of negative reproductive events ^d				
Yes	19 (27.1)	43 (33.6)	130 (34.1)	217 (36.7)
No	51 (72.9)	83 (64.8)	246 (64.6)	365 (61.7)
Somatic comorbidity (MoBa Q1)				
No	52 (74.4)	99 (77.3)	295 (77.4)	442 (74.7)
Medicated	9 (12.8)	15 (11.7)	38 (10.0)	85 (14.4)
Non-medicated	9 (12.8)	14 (10.9)	48 (12.6)	65 (10.9)
Mental comorbidity (MoBa Q1)				
No	61 (87.1)	120 (93.8)	327 (85.8)	514 (86.8)
Medicated	4 (5.7)	4 (3.1)	21 (5.5)	23 (4.0)
Non-medicated	5 (7.1)	4 (3.1)	33 (8.7)	55 (9.2)

Notes: ^aWomen's income status (USD/year): 1.0 NOK≈0.13 USD. ^bSmokers consist of daily and occasional smokers. ^cIncludes both major depression with and without negative lifetime event. ^dPrevious negative event includes spontaneous abortion/stillbirth, ectopic pregnancy, and termination of pregnancy (MoBa Q1).

Abbreviations: BMI, body mass index; LTHMD, lifetime history of major depression; MoBa Q1, Norwegian Mother and Child Cohort Study questionnaire 1; NOK, Norwegian Kroner.

Table 4 Drug utilization in the adherence groups

	Adherence groups			
Drug utilization (mean $\pm$ SE)	Decreasing-Low, n=70 (5.8%)	Increasing-Medium, n=128 (11.0%)	Constant-Medium, n=381 (32.9%)	Constant-High, n=592 (50.2%)
DDD during gestation	55 $\pm$ 5.62	112 $\pm$ 6.09	129 $\pm$ 3.64	251 $\pm$ 3.66
PDC ^a during gestation	0.20 $\pm$ 0.02	0.40 $\pm$ 0.02	0.45 $\pm$ 0.01	0.82 $\pm$ 0.01
Medication gaps ^b during gestation	0.71 $\pm$ 0.10	0.88 $\pm$ 0.09	1.28 $\pm$ 0.06	1.01 $\pm$ 0.04
Prescriptions during gestation	1.60 $\pm$ 0.13	2.64 $\pm$ 0.09	2.55 $\pm$ 0.05	3.13 $\pm$ 0.04
Prescriptions 6 months prior to LMP	3.44 $\pm$ 0.50	1.18 $\pm$ 0.21	5.95 $\pm$ 0.19	7.80 $\pm$ 0.18
DDD 6 months prior to LMP	34 $\pm$ 5.49	4 $\pm$ 0.88	78 $\pm$ 2.41	139 $\pm$ 2.55

Notes: ^aPDC, total sum of DDDs during gestation/length of gestation in days. ^bMedication gap is break of 14 days between the previous and next dispensed prescription.

Abbreviations: DDD, defined daily dose; LMP, last menstrual period; PDC, proportion of days covered; SE, standard error.

Table 5 Multivariate crude and adjusted OR for Decreasing-Low (n=70)^a vs other THRT trajectory (n=1,101)

Variables	Crude OR (95% CI)	Adjusted ^b OR (95% CI)
Educational level-ongoing (years)		
<9	4.31 (1.08–17.28)	4.10 (1.05–16.03)
9–12	1	1
12–16	0.74 (0.39–1.38)	0.72 (0.38–1.35)
>16	0.72 (0.36–1.41)	0.70 (0.35–1.38)
History of negative reproductive events		
Yes	0.66 (0.38–1.14)	0.66 (0.38–1.14)
No	1	1

Notes: ^aThere was missing covariates information in 27.7% of the hypothyroid group (n=1,171): education (1.7%), smoking (20.3%), LTHMD (1.5%), BMI (2.6%), income (2.8%), pregnancy planning (0.9%), and negative reproductive history (1.5%).

^bIndependent variables included in the main GEE model were educational level and history of negative reproductive events. These variables were selected using a combined, forward- and backward-model selection, based on the lowest AIC value.

Abbreviations: AIC, Akaike Information Criterion; BMI, body mass index; GEE, generalized estimating equation; LTHMD, lifetime history of major depression; THRT, thyroid hormone replacement therapy.

The agreement and sensitivity in the second trimester might be affected by the weeks 13–16. The MoBa questionnaires overlap for the assessment of exposure around weeks 13–16 but with different questions (“thyroid” treatment in MoBa Q1 vs “other medications” in MoBa Q3). Hence, women might have reported THRT use more correctly in MoBa Q1, given the indication-specific rather than open-ended question. Due to the overlap for time period 13–16 weeks, we aggregated both intervals, and women were exposed if they crossed out this time period in MoBa Q1 or MoBa Q3. This could have led to higher agreement for this month (13–16 weeks) and the second trimester, and such possibility should be kept in mind, when interpreting the results. Besides the questionnaire design and length of time intervals, there might exist other explanations for the observed drop in self-reporting in late gestation, such as retrospective reporting and negative pregnancy outcome. For example, after a miscarriage or stillbirth, there might be a higher probability of dropout or non-completion of the late

pregnancy questionnaire.²⁸ This results in a lack of consistency between THRT estimates based on MoBa data and those from other data sources, such as NorPD.

Our findings are generally similar to those of three prior studies, albeit with some differences in sensitivity (84.5% in our study vs 70.0%–85.6% in the other studies).^{21–23} Though our results vary slightly from previous studies,^{21–23} the conclusions are consistent, and the slight differences might stem from different study designs, reference standards, as well as variations in time elapsed between response to questionnaires (self-report) and actual exposure time.^{21–23}

In addition, around 2% of women in the study population used THRT. This result is consistent with the Norwegian study by Bjoro et al,²⁹ which estimated that around 1%–2% of women in their reproductive age used thyroxine.²⁹ A higher proportion of THRT exposure during pregnancy was found in the French study by Demailly et al.³⁰ Geographic and ethnic variations in prevalence of thyroid disorders, and possible differences in definition of thyroid disease and severity, or in treatment guidelines and practice might explain this discrepancy.^{31,32}

Based on agreement between data sources and our background knowledge about the MoBa questionnaires, we specified NorPD as a reference standard in our validation. A possible caveat is that the validity of self-reports on THRT might be underestimated.²⁸ For safety studies that rely on mid and late gestational exposure of THRT, NorPD seems to be a better source for exposure definition, as MoBa, which by design, did not capture THRT in the late second and third trimesters.

When exploring adherence patterns with GBTM among women who self-reported and filled a prescription for THRT, we identified four disjoint groups.

None of the identified trajectories followed the shape of a very low and stable trajectory, or a trajectory that decreased from very high or medium to medium or low

adherence during pregnancy. In addition, there was no evident increase of THRT adherence after delivery relative to the pregnancy period. This is contrary to prior studies on antidepressant or statin use during pregnancy, where fear of teratogenic effects due to drug exposure can probably explain the constantly low or decreasing adherence during pregnancy.^{8,33} Additionally, adherence kept decreasing even after delivery in the D-L group. It is possible that women in the D-L group might have less severe conditions, such as subclinical hypothyroidism.

Low adherence among 6% of women with hypothyroidism might also be explained by the fact that THRT, such as levothyroxine, is still perceived as risky by pregnant, lactating, or women attempting to conceive, even though its usage is safe and recommended when clinically needed.^{34,35} However, based on the results presented, pregnancy did not seem to be a major driver of low medication adherence in the context of THRT.

Women in the C-M group would benefit from medical surveillance of THRT use during pregnancy, as they are characterized by frequent gaps between dispensing prescriptions. Having a LTHMD was a common characteristic of women in the C-M and D-L groups. This suggests that poorer mental health may negatively affect a spectrum of women's health behaviors, including treatment of nonpsychiatric disorders.³⁶ The high rate of obesity and negative reproductive events among women in the C-H group may indicate more severe hypothyroidism.^{37,38} It is also interesting that the stable dynamics are those associated with relative high adherence to THRT. This underpins the possibility of more severe hypothyroid conditions associated with women within these groups, who might keep THRT constant to ensure well-being.

While GBTMs have been used as measures of adherence for other medication such as statins, this is the first time longitudinal trajectory models have been applied to THRT.⁸ For THRT, the PDC and the DDD levels also matched the trajectory dynamics of the four adherence groups. In summary, we obtained a better understanding of long-term drug utilization around gestation and identified characteristics which are linked to certain adherence patterns.

Differences in parity, smoking habits, and BMI in the hypothyroid group compared to a reference group have been confirmed in the study by Männistö et al.³⁹ Our finding that the hypothyroid population had higher percentage of LTHMD was confirmed by literature that reported a link between thyroid disorders and depression.⁴⁰ The regression analysis showed that women with less education are more likely to belong to the D-L group. Lower educational level may have

negative impact on understanding concepts of medication use, such as side effects or dosage schedule.⁴¹ Less knowledge about THRT and its use during pregnancy might explain why women in this group use treatment more sporadically. Fear about adverse side effects of THRT for the unborn child cannot be the only explanation as there is also a decrease in prescriptions after delivery. For women with low education level, prepregnancy counseling about accurate use of THRT might improve adherence in this group.⁴¹ More research is however needed to identify characteristics of women in the low THRT adherence group, and reasons for low adherence, in order to optimize intervention strategies.

The use of high-quality and multiple data sources was a clear strength of this study.^{11,15} The advantage of using NorPD as reference standard is that the information is objective and satisfactorily valid for chronic medication treatment. In comparison to other data sources (eg, self-report) on long-term THRT, NorPD represents a more complete, valid, and reliable data source.¹⁵ In addition, prescription records are not affected by recall bias. Using a relatively novel methodological approach to identify different adherence patterns allowed us to gain a more complete picture about long-term adherence in the time around pregnancy.⁴² Furthermore, with the specification of the research question on THRT, we showed that agreement was consistent with medication use for other chronic disorders and presented consistent percentages of THRT use during pregnancy in Norway.^{6,29} Exploring patterns of adherence to THRT across pregnancy may inform future research about the importance of evaluating the consequences of suboptimal medication adherence on maternal-child health. Not least, knowledge about characteristics of less compliant women might be of value for adherence interventions.

An important study limitation is selection bias in the MoBa data.¹³ It is known that women participating in MoBa are on average older and have a better socioeconomic situation and smoke less often during pregnancy than the overall Norwegian population.¹³ Since the hypothyroid population reflected this selection process with respect to age, socioeconomic status, and smoking, the identified adherence group percentages might not be representative of the overall hypothyroid pregnant population in Norway. While the true adherence group percentages might be higher for the C-H group, the identified patterns of THRT adherence are only approximations to the real use of THRT around gestation.⁴³⁻⁴⁵ We could however validate that the majority of patients continue THRT in the long term once started, as well as specific characteristics of hypothyroid population.^{39,46}

Another limitation is that the MoBa questions about thyroid disease and treatment are framed differently in the beginning compared to the end of pregnancy. This can result in underestimation of the degree of THRT use in mid to late gestational periods. NorPD does not reflect actual medication use, only what has been dispensed from the pharmacy. Women could decide not to take the medication or take it differently than prescribed.^{21,47} Furthermore, NorPD does not capture medications bought online through other channels than Norwegian pharmacies or treatment given during hospital stays.²¹ In total, NorPD might overestimate prenatal drug exposure, as well as underestimate validity of self-report.²⁸

The lower agreement in the 4-week intervals could stem from discrepancies in dates of supply and maternal report of THRT use. Indeed, such a granular window of exposure is more susceptible to inaccuracy than broader windows like the trimesters of pregnancy.

Although we do not consider the discrepancy in dates of supply to have major impact on our results, we cannot completely rule out some influence.

We have no information about the severity of hypothyroidism. Disease severity however might be an important determinant of THRT use and adherence to prescribed treatment. Future studies should include this information.

The fact that we had almost 30% missing information on important variables is a data limitation. However, to limit this risk of bias, we carried out multiple imputation. A drawback of this analysis was that, also after multiple imputation, the cell counts for some independent variables in the D-L group remained low, that is <5, leading to less precise effect estimates. We cannot rule out that the cell count also affected the choice of selected variables in the model selection process. These limitations of the analysis present a caveat to the reliability of these results.

Conclusion

Women with hypothyroidism tend to accurately report use of prescribed THRT during early pregnancy. Differences in exposure measurement in the MoBa questionnaires in the beginning and toward the end of pregnancy can likely influence the accuracy of recall, and thus the agreement between the two data sources, especially in mid to late gestational periods.

Given that approximately one in two women adhered adequately to THRT around pregnancy, there is still room for improvement. Behavioral and educational interventions targeted to women, especially those with unfavorable socioeconomic characteristics, could improve adherence to

THRT. Likewise, women with hypothyroidism should be empowered to develop an evidence-based understanding about the importance of THRT in the perinatal period, to safeguard maternal and child health. Using GBTM, this study illustrated how feasible it is to identify attributes, which are linked to medication adherence and long-term medication use.

Abbreviations

ATC, Anatomical Therapeutic Chemical
BMI, body mass index
C-H, Constant-High
C-M, Constant-Medium
DDD, defined daily dose
D-L, Decreasing-Low
GBTM, group-based trajectory model
I-M, Increasing-Medium
LTHMD, lifetime history of major depression
MBRN, Medical Birth Registry of Norway
MoBa, Norwegian Mother and Child Cohort Study
MoBa Q1, MoBa questionnaire 1
MoBa Q3, MoBa questionnaire 3
MoBa Q4, MoBa questionnaire 4
NorPD, Norwegian Prescription Database
NPV, negative predictive value
PDC, proportion of days covered
PPV, positive predictive value
THRT, thyroid hormone replacement therapy

Acknowledgments

The authors are grateful to all of the participating families in Norway who took part in this ongoing cohort study. The Norwegian Mother and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and by the Ministry of Education and Research, National Institutes of Health (NIH)/National Institute of Neurological Disorders and Stroke (NINDS) (grant no. 1 UO1 NS 047537-01 and grant no. 2 UO1 NS 047537-06A1). This project and Anna S Frank's PhD research fellowship are funded by the Norwegian Women's Public Health Association. Angela Lupattelli and Hedvig Nordeng are funded by the H2020 European Research Council Starting Grant, "DrugsInPregnancy" (grant no. 639377). Anna S Frank's research visit at Cornell University was made possible by a Kristine Bonnevie travel stipend and by a NORBIS international travel grant. David S Matteson received financial support from a Xerox PARC Faculty Research Award, National Science Foundation grant DMS-1455172,

Cornell University Institute of Biotechnology and the New York State Division of Science, Technology and Innovation (NYSTAR), and Cornell University Atkinson's Center for a Sustainable Future AVF-2017.

Author contributions

All authors contributed to data analysis, drafting and revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure

HN is a member of several nonprofit organizations. She is a board member of the Norwegian Pharmaceutical Society, member of the scientific board of the European Network of Teratology Information Services, chair of the Pregnancy Special Interest Group (SIG) and International Society of Pharmacoeconomics (ISPE), and member of the Executive Committee, European Drug Utilization Group (EuroDURG). She serves as an independent expert and a member of the Pharmacovigilance Risk Assessment Committee (PRAC) and European Medicines Agency (EMA). AL is the current head of Steering Committee of the Norwegian Society for Pharmacoeconomics, which is a nonprofit organization for pharmacoeconomists in Norway. ASF, AL, and HN are part of the Pharmacoeconomics and Drug Safety Research Group (PharmaSafe). PharmaSafe is a member of the European Network of Centers for Pharmacoeconomics and Pharmacovigilance (EnCePP). DSM reports no conflicts of interest in this work.

References

- Klubo-Gwiezdzińska J, Burman KD, Van Nostrand D, Wartofsky L. Levothyroxine treatment in pregnancy: indications, efficacy, and therapeutic regimen. *J Thyroid Res*. 2011;2011:843591.
- Abalovich M, Gutierrez S, Alcaraz G, Maccallini G, Garcia A, Levalle O. Overt and subclinical hypothyroidism complicating pregnancy. *Thyroid*. 2002;12(1):63–68.
- Liu X, Andersen SL, Olsen J, et al. Maternal hypothyroidism in the perinatal period and childhood asthma in the offspring. *Allergy*. 2018;73(4):932–939.
- West SL, Ritchey ME, Poole C. Validity of pharmacoeconomic drug and diagnosis data. In: Strom BL, Kimmel SE, Hennessy S, editors. *Pharmacoeconomics*. 5th ed. New Jersey: Wiley-Blackwell; 2012:757–794.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
- Palmsten K, Hulugalle A, Bandoli G, et al. Agreement between maternal report and medical records during pregnancy: medications for rheumatoid arthritis and asthma. *Paediatr Perinat Epidemiol*. 2018;32(1):68–77.
- Shrank WH, Hoang T, Ettner SL, et al. The implications of choice: prescribing generic or preferred pharmaceuticals improves medication adherence for chronic conditions. *Arch Intern Med*. 2006;166(3):332–337.
- Franklin JM, Shrank WH, Pakes J, et al. Group-based trajectory models: a new approach to classifying and predicting long-term medication adherence. *Med Care*. 2013;51(9):789–796.
- Dillon P, Stewart D, Smith SM, Gallagher P, Cousins G. Group-based trajectory models: assessing adherence to antihypertensive medication in older adults in a community pharmacy setting. *Clin Pharmacol Ther*. 2018;103(6):1052–1060.
- Nagin DS. Group-based trajectory modeling: an overview. *Ann Nutr Metab*. 2014;65(2–3):205–210.
- Magnus P, Birke C, Vejrup K, et al. Cohort profile update: the norwegian mother and child cohort study (MoBa). *Int J Epidemiol*. 2016;45(2):382–388.
- Magnus P, Irgens LM, Haug K, et al; MoBa Study Group. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol*. 2006;35(5):1146–1150.
- Nilsen RM, Vollset SE, Gjessing HK, et al. Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatr Perinat Epidemiol*. 2009;23(6):597–608.
- Irgens LM. The Medical Birth Registry of Norway. Epidemiological research and surveillance throughout 30 years. *Acta Obstet Gynecol Scand*. 2000;79(6):435–439.
- Furu K. Establishment of the nationwide Norwegian Prescription Database (NorPD)-new opportunities for research in pharmacoepidemiology in Norway. *Norsk epidemiologi*. 2008;18(2):129–136.
- Norwegian Mother and Child Cohort Study (MoBa) - Questionnaires from MoBa. 2005. Available from: <https://www.fhi.no/en/studies/moba/for-forskere-artikler/questionnaires-from-moba/>. Accessed January 15, 2017.
- World Health Organization Collaborating Centre for Drug Statistics Methodology. ATC/DDD index 2018. 2017. Available from: https://www.whocc.no/atc_ddd_index/. Accessed January 20, 2017.
- Hansen RA, Dusetzina SB, Dominik RC, Gaynes BN. Prescription refill records as a screening tool to identify antidepressant non-adherence. *Pharmacoeconomics Drug Saf*. 2010;19(1):33–37.
- Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. The lifetime history of major depression in women. Reliability of diagnosis and heritability. *Arch Gen Psychiatry*. 1993;50(11):863–870.
- de Jong van den Berg LT, Feenstra N, Sorensen HT, Cornel MC, Toft Sorensen H, Group E. Improvement of drug exposure data in a registration of congenital anomalies. Pilot-study: pharmacist and mother as sources for drug exposure data during pregnancy. EuroMAP Group. European Medicine and Pregnancy Group. *Teratology*. 1999;60(1):33–36.
- Sarangarm P, Young B, Rayburn W, et al. Agreement between self-report and prescription data in medical records for pregnant women. *Birth Defects Res A Clin Mol Teratol*. 2012;94(3):153–161.
- van Gelder M, Vorstenbosch S, Te Winkel B, van Puijenbroek EP, Roeleveld N. Using web-based questionnaires to assess medication use during pregnancy: a validation study in 2 prospectively enrolled cohorts. *Am J Epidemiol*. 2018;187(2):326–336.
- Pisa FE, Casetta A, Clagnan E, Michelesio E, Vecchi Brumatti L, Barbone F. Medication use during pregnancy, gestational age and date of delivery: agreement between maternal self-reports and health database information in a cohort. *BMC Pregnancy Childbirth*. 2015;15:310.
- Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol*. 2010;6:109–138.
- Jones BL, Nagin DS. A Stata plugin for estimating group-based trajectory models. 2012. Available from: https://ssrc.indiana.edu/doc/wimdocs/2013-03-29_nagin_trajectory_stata-plugin-info.pdf. Accessed November 22, 2017.
- van Buuren S, Groothuis-Oudshoorn K, Robitzsch A, et al. Package 'mice': Multivariate Imputation by Chained Equations. R Package Version 3.3.0. 2018. Available from: <http://stefvanbuuren.github.io/mice/>. Accessed July 20, 2018.
- Lumley T. Package 'survey': Analysis of Complex Survey Samples. R Package Version 3.33-2. 2018. Available from: <http://r-survey.r-forge.r-project.org/survey/>. Accessed July 24, 2018.
- van Gelder MM, van Rooij IA, de Walle HE, Roeleveld N, Bakker MK. Maternal recall of prescription medication use during pregnancy using a paper-based questionnaire: a validation study in the Netherlands. *Drug Saf*. 2013;36(1):43–54.

29. Bjoro T, Holmen J, Krüger O, et al. Prevalence of thyroid disease, thyroid dysfunction and thyroid peroxidase antibodies in a large, unselected population. The Health Study of Nord-Trøndelag (HUNT). *Eur J Endocrinol*. 2000;143(5):639–647.
30. Demailly R, Escolano S, Quantin C, Tubert-Bitter P, Ahmed I. Prescription drug use during pregnancy in France: a study from the national health insurance permanent sample. *Pharmacoepidemiol Drug Saf*. 2017;26(9):1126–1134.
31. Garmendia Madariaga A, Santos Palacios S, Guillén-Grima F, Galofré JC. The incidence and prevalence of thyroid dysfunction in Europe: a meta-analysis. *J Clin Endocrinol Metab*. 2014;99(3):923–931.
32. Saraladevi R, Nirmala Kumari T, Shreen B, Rani U V. Prevalence of thyroid disorder in pregnancy and pregnancy outcome. *LAJM*. 2016;3(3): 1–11.
33. Hurault-Delarue C, Chouquet C, Savy N, et al. How to take into account exposure to drugs over time in pharmacoepidemiology studies of pregnant women? *Pharmacoepidemiol Drug Saf*. 2016;25(7):770–777.
34. Petersen I, McCrea RL, Lupattelli A, Nordeng H. Women's perception of risks of adverse fetal pregnancy outcomes: a large-scale multinational survey. *BMJ Open*. 2015;5(6):e007390.
35. Alexander EK, Pearce EN, Brent GA, et al. 2017 Guidelines of the american thyroid association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*. 2017;27(3):315–389.
36. Dimatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. *Arch Intern Med*. 2000;160(14):2101–2107.
37. Sanyal D, Raychaudhuri M. Hypothyroidism and obesity: An intriguing link. *Indian J Endocrinol Metab*. 2016;20(4):554–557.
38. Smith GC, Fretts RC. Stillbirth. *Lancet*. 2007;370(9600):1715–1725.
39. Männistö T, Vääräsmäki M, Pouta A, et al. Perinatal outcome of children born to mothers with thyroid dysfunction or antibodies: a prospective population-based cohort study. *J Clin Endocrinol Metab*. 2009;94(3):772–779.
40. Hage MP, Azar ST. The link between thyroid function and depression. *J Thyroid Res*. 2012;2012:590648.
41. Alkatheri AM, Albekairy AM. Does the patients' educational level and previous counseling affect their medication knowledge? *Ann Thorac Med*. 2013;8(2):105–108.
42. Hess LM, Raebel MA, Conner DA, Malone DC. Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. *Ann Pharmacother*. 2006;40(7–8):1280–1288.
43. Kandukuri RC, Khan MA, Soltys SM. Nonadherence to medication in hypothyroidism: a case report. *Prim Care Companion J Clin Psychiatry*. 2010;12(3):PCC.09m00863.
44. Briesacher BA, Andrade SE, Fouayzi H, Chan KA. Comparison of drug adherence rates among patients with seven different medical conditions. *Pharmacotherapy*. 2008;28(4):437–443.
45. Nagin DS, Tremblay RE. What has been learned from group-based trajectory modeling? Examples from physical aggression and other problem behaviors. *Ann Am Acad Pol Soc Sci*. 2005;602(1):82–117.
46. Rodriguez-Gutierrez R, Maraka S, Ospina NS, Montori VM, Brito JP. Levothyroxine overuse: time for an about face? *Lancet Diabetes Endocrinol*. 2017;5(4):246–248.
47. Cohen JM, Wood ME, Hernandez-Diaz S, Nordeng H. Agreement between paternal self-reported medication use and records from a national prescription database. *Pharmacoepidemiol Drug Saf*. 2018;27(4):413–421.

Supplementary materials

Additional details on multiple imputation analysis

We assumed that the data were missing at random and that the missing information depended on the observed variables. We therefore imputed the missing information using multiple imputation by chained equations.¹ We used a Gibbs sampler to draw from conditional distributions that were generated from an initial random imputation of all missing variables.^{1,2}

In our analysis, we created 10 imputed data sets and let the Gibbs sampler run for 120 iterations to ensure convergence. This number of iterations is enough as each imputation set depends on the previous imputed variables, together with the other auxiliary variables.¹ The auxiliary variables for imputing the missing information were age, body mass index, marital status, pregnancy planning, income, educational level,

smoking, both comorbidities, recommended supplement use, parity, history of negative reproductive event, LTHMD, and the binary outcome variable. In addition, we included variables which might influence nonresponse, such as fetal survival, indicators of plural pregnancies (twin/triplets), and prescription records and self-report indicator variables. Furthermore, we added the variable that characterizes the four trajectory groups as well as the variable presenting the percentage of days covered.

Before the final regression analysis, predictor variables were chosen using combined forward and backward stepwise-model selection on each of the 10 imputed data sets. The model with the lowest Akaike Information Criterion value for each imputed set was chosen. Finally, if a variable was selected at least on one of the 10 completed sets, we included it in the final model.

Table S1 Calculation of sensitivity, specificity, PPV, and NPV^a

	NorPD reference standard		
	Positive (truly exposed)	Negative (truly unexposed)	
MoBa questionnaire positive (exposed)	A ^b ("Both")	B ^b ("MoBa only")	"MoBa total"
MoBa questionnaire negative (unexposed)	C ^b ("NorPD only")	D ^b ("Neither")	"Total"–"MoBa total"
	"NorPD total"	"Total"–"NorPD total"	"Total" ^c

Notes: ^aSensitivity: $A/(A+C)*100$, sensitivity standard error $(SE_sensitivity)=((sensitivity*(1-sensitivity))/(A+C))^{1/2}$; specificity: $D/(D+B)*100$, specificity standard error $(SE_specificity)=((specificity*(1-specificity))/(D+B))^{1/2}$; PPV: $A/(A+B)*100$, PPV standard error $(SE_PPV)=((PPV*(1-PPV))/(A+B))^{1/2}$; NPV: $D/(D+C)*100$, NPV standard error $(SE_NPV)=((NPV*(1-NPV))/(D+C))^{1/2}$; given the standard error (SE), we calculate 95% CI, with Z-score (Z) of 1.96 according to $(x-Z*SE, x+Z*SE)$, with x representing sensitivity, specificity, PPV, or NPV and SE the respective standard errors. ^bTable cell counts correspond to the number of consistent (ie 'Both', 'Neither') and inconsistent (ie 'MoBa only', 'NorPD only') self-reports and prescription records, and depend on THRT exposure length. ^cTotal study population, n=56,148.

Abbreviations: MoBa, Norwegian Mother and Child Cohort Study; NorPD, Norwegian Prescription Database; NPV, negative predictive value; PPV, positive predictive value; THRT, thyroid hormone replacement therapy.

Table S2 Agreement between MoBa and NorPD for 4-week intervals in MoBa Q1 and MoBa Q3

THRT		MoBa total ^a		NorPD total ^a		MoBa only	NorPD only	Both	Neither	Agreement	
		n	%	n	%	n	n	n	n	κ^b (95% CI)	Range ^c
Q1	Weeks 0–4	959	1.7	923	1.6	231	202	721	54,987	0.76 (0.76–0.77)	S
	Weeks 5–8	968	1.7	941	1.7	238	211	730	54,969	0.76 (0.75–0.77)	S
	Weeks 9–12	978	1.7	960	1.7	232	214	746	54,956	0.77 (0.76–0.77)	S
Q3	Weeks 13–16 ^d	1,065	1.9	957	1.7	262	154	803	54,929	0.79 (0.78–0.80)	S
	Weeks 17–20	330	0.6	985	1.7	68	723	262	55,095	0.39 (0.38–0.40)	f
	Weeks 21–24	334	0.6	1,031	1.8	69	766	265	55,048	0.38 (0.37–0.39)	f
	Weeks 25–28	338	0.6	1,042	1.8	66	770	272	55,040	0.39 (0.38–0.40)	f

Notes: ^aTotal study population, n=56,148. ^bCohen's kappa coefficient (κ). ^cSubdivision of agreement by Landis and Koch: <0, no (no); 0–0.20, slight (s); 0.21–0.40, fair (f); 0.41–0.60, moderate (m); 0.61–0.80, substantial (S); and 0.81–1.00, perfect (P). ^dWeeks 13–16 aggregate the text box "week 13+" from MoBa Q1 and "weeks 13–16" from MoBa Q3.

Abbreviations: MoBa, Norwegian Mother and Child Cohort Study; MoBa Q1, MoBa questionnaire 1; MoBa Q3, MoBa questionnaire 3; NorPD, Norwegian Prescription Database; THRT, thyroid hormone replacement therapy.

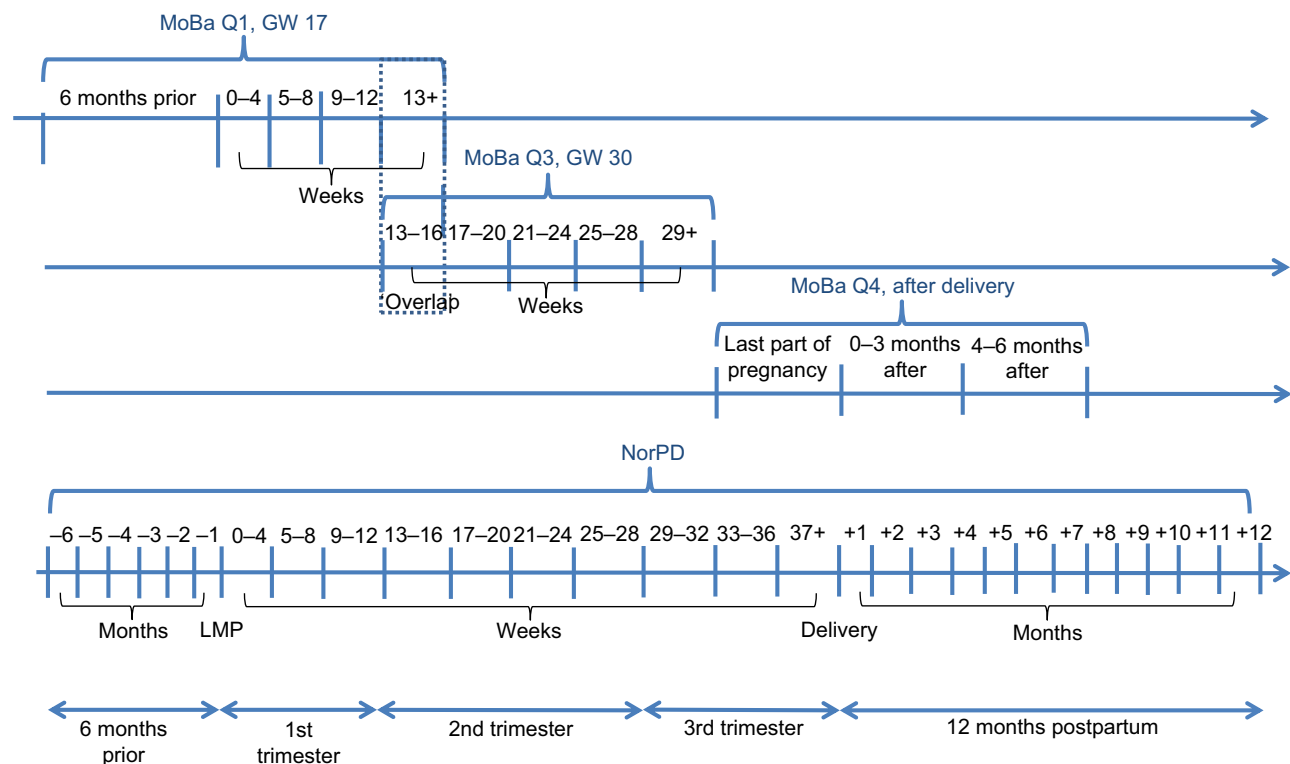
Table S3 Model selection using BIC criterion, and estimated group proportions

Group numbers	BIC ^b	Polynomial order of each group trajectory ^c	Estimated group proportions ^a				
			Group 1 (%)	Group 2 (%)	Group 3 (%)	Group 4 (%)	Group 5 (%)
1	-18,540	2	100	–	–	–	–
2	-16,263	2, 3	39.0	61.0	–	–	–
3	-15,939	2, 1, 3	5.8	37.7	56.5	–	–
4	-15,640	2, 3, 0, 3	5.8	11.0	33.0	50.2	–
5	-15,539	1, 3, 3, 2, 3	4.7	19.3	9.2	31.5	35.3

Notes: ^aEstimated group proportions should exceed 5.0%. ^bBIC (for the total number of participants) was calculated as $BIC = 2 \cdot \log(L) - k \cdot \log(N)$, where L is the maximal value of the model's likelihood function, k represents the number of model parameter, and N the sample size. ^cThe model with the highest (least negative) BIC value is preferred.

^cTrajectory shapes: 0, constant; 1, linear; 2, quadratic; and 3, cubic.

Abbreviation: BIC, Bayesian Information Criterion.

**Figure S1** Timeline of questionnaires (MoBa) and of prescription records (NorPD).

Notes: MoBa Q1, completed in gestational week 17, covers medication use in early pregnancy. MoBa Q2 is a dietary questionnaire filled out around gestational week 22 and was not included in this study. MoBa Q3, completed in gestational week 30, covers medication use in mid pregnancy. MoBa Q4, completed 6 months after delivery, covers medication use in last part of pregnancy and during the first 6 months after delivery. We did not include the first 6 months after delivery covered by MoBa Q4 in our analysis.

Abbreviations: GW, gestational week; LMP, last menstrual period; MoBa, Norwegian Mother and Child Cohort Study; MoBa Q1, MoBa questionnaire 1; MoBa Q2, MoBa questionnaire 2; MoBa Q3, MoBa questionnaire 3; MoBa Q4, MoBa questionnaire 4; NorPD, Norwegian Prescription Database.

References

1. van Buuren S, Groothuis-Oudshoorn K, Robitzsch A, et al. Package “mice”: Multivariate Imputation by Chained Equations. R Package Version 3.3.0. 2018. Available from: <http://stefvanbuuren.github.io/mice/>. Accessed July 20, 2018.
2. Tarantola A. Monte Carlo methods. In: *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia: SIAM; 2005:41–55.
3. Fraley C, Raftery AE. How many clusters? Which clustering method? Answers via model-based cluster analysis. *Comput J*. 1998;41(8): 578–588.

Clinical Epidemiology**Dovepress****Publish your work in this journal**

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>