



Calculating daily dose in the Observational Medical Outcomes Partnership Common Data Model

Theresa Burkard¹  | Kim López-Güell¹ | Artem Gorbachev² | Lucía Bellas³ | Annika M. Jödicke¹ | Edward Burn¹ | Maria de Ridder⁴ | Mees Mosseveld⁴ | Jasmine Gratton⁵ | Sarah Seager⁵ | Dina Vojinovic⁶ | Miguel Angel Mayer^{7,8} | Juan Manuel Ramírez-Anguita^{8,9} | Angela Leis Machín⁸ | Marek Oja¹⁰ | Raivo Kolde¹⁰ | Klaus Bonadt¹¹ | Daniel Prieto-Alhambra^{1,4}  | Chistian Reich² | Martí Català¹

¹Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK

²Odysseus Data Services, Cambridge, Massachusetts, USA

³Clinical Pharmacology Service, Hospital Universitari Vall d'Hebron, Barcelona, Spain

⁴Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands

⁵IQVIA Ltd, London, UK

⁶IQVIA Solutions B.V, Amsterdam, the Netherlands

⁷Management and Control Department, Hospital del Mar Barcelona, Barcelona, Spain

⁸Research Program on Biomedical Informatics (GRIB), Hospital del Mar Research Institute (IMIM), Barcelona, Spain

⁹Universitat Pompeu Fabra, Barcelona, Spain

¹⁰Institute of Computer Science, University of Tartu, Tartu, Estonia

¹¹IQVIA Commercial GmbH & Co. OHG, Frankfurt, Germany

Correspondence

Daniel Prieto-Alhambra, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Windmill Road, OX37LD, Oxford, UK.

Email: daniel.prietoalhambra@ndorms.ox.ac.uk

Funding information

Darwin EU, Grant/Award Number: EMA/2021/08/TDA; EHDEN

Abstract

Purpose: We aimed to develop a standardized method to calculate daily dose (i.e., the amount of drug a patient was exposed to per day) of any drug on a global scale using only drug information of typical observational data in the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) and a single reference table from Observational Health Data Sciences And Informatics (OHDSI).

Materials and Methods: The OMOP DRUG_STRENGTH reference table contains information on the strength or concentration of drugs, whereas the OMOP DRUG_EXPOSURE table contains information on patients' drug prescriptions or dispensations/claims. Based on DRUG_EXPOSURE data from the primary care databases Clinical Practice Research Datalink GOLD (United Kingdom) and Integrated Primary Care Information (IPCI, The Netherlands) and healthcare claims from PharMetrics® Plus for Academics (USA), we developed four formulas to calculate daily dose given

Chistian Reich and Martí Català contributed equally.

Preliminary results of this work have been presented at the Observational Health Data Sciences and Informatics (OHDSI) Symposium held in East Brunswick, USA, on October 20–22, 2023.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Pharmacoepidemiology and Drug Safety* published by John Wiley & Sons Ltd.

different DRUG_STRENGTH reference table information. We tested the dose formulas by comparing the calculated median daily dose to the World Health Organization (WHO) Defined Daily Dose (DDD) for six different ingredients in those three databases and additional four international databases representing a variety of healthcare settings: MAITT (Estonia, healthcare claims and discharge summaries), IQVIA Disease Analyzer Germany (outpatient data), IQVIA Longitudinal Patient Database Belgium (outpatient data), and IMASIS Parc Salut (Spain, hospital data). Finally, in each database, we assessed the proportion of drug records for which daily dose calculations were possible using the suggested formulas.

Results: Applying the dose formulas, we obtained median daily doses that generally matched the WHO DDD definitions. Our dose formulas were applicable to >85% of drug records in all but one of the assessed databases.

Conclusion: We have established and implemented a standardized daily dose calculation in OMOP CDM providing reliable and reproducible results.

KEYWORDS

common data model, dose, OMOP, WHO DDD

Key Points

- Each database in the large Observational Health Data Sciences And Informatics (OHDSI) research network having their data in the OMOP CDM can now apply a standardized way to calculate drug daily dose.
- Only the OMOP DRUG_STRENGTH reference table and DRUG_EXPOSURE table are needed for the dose formulas.
- Four different dose formulas apply to the majority of drug records.
- Most of the calculated median daily dose matched the World Health Organization (WHO) Defined Daily Dose (DDD).
- The established dose formulas increase transparency, reliability, and reproducibility of daily dose-related research in real-world evidence.

Plain Language Summary

The analysis of previously collected healthcare data is crucial for understanding the effects of drugs. However, calculating the amount of a drug ingredient a patient was exposed to per day has been a challenge: (i) Such calculations are different for different drug types, for example, tablets, and liquid drugs; and (ii) it requires a comprehensive source of information about how much drug ingredient is in each drug product on the market. In this research, we propose a solution to this problem. We introduced a standardized method for calculating the daily dose for all possible drugs in healthcare data. Moreover, by working our way through the available anonymized patient data in a structured way, we made the dose calculation process transparent and reproducible. Furthermore, comparing our method against global standards according to the World Health Organization (WHO) average dose per day yielded overall good results. This research offers a clear roadmap for increasing reliability and reproducibility of daily dose results in research using healthcare data. Ultimately, patients will benefit from an improved process of generating evidence of drugs related to different quantities of exposure.

1 | BACKGROUND

In the realm of pharmacoepidemiology and drug safety, the analysis of healthcare data plays a pivotal role in shaping our understanding of the real-world effects of drugs. To unlock the full potential of those

data, robust standardized and well-defined patient-level information are needed. The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) in combination with the Standardized Vocabularies¹ has emerged as a cornerstone in providing a common structure for analyzing healthcare data.² It provides a standardized

way to represent clinical and healthcare concepts, ensuring consistency and interoperability across diverse data sources. The Observational Health Data Sciences and Informatics (OHDSI) community, which maintains the OMOP CDM, has produced reliable and transparent analytic packages, facilitating large-scale analysis of observational healthcare databases.³ However, conducting drug utilization studies in the context of the European Union's Data Analysis and Real-World Interrogation Network (DARWIN EU)⁴ initiative revealed that a standardized way to calculate drug daily dose was still missing.

In general, daily dose is not readily available from the data. This becomes even more challenging when faced with international multi-center and multi-database studies as drug products differ between countries and so do drug vocabularies. OMOP overcomes this challenge by providing the OHDSI Standardized Vocabularies in which drug products are represented through RxNorm and RxNorm Extension for products outside the USA. In our previous paper we describe how this reference came about.¹ The DRUG_STRENGTH reference table⁵ stores information concerning strength or concentration of a drug and the respective units. The DRUG_EXPOSURE table⁶ stores all drug records and related information (e.g., product, duration, quantity) that were prescribed, dispensed, or administered to patients. Thus, DRUG_STRENGTH together with prescription and administration data in DRUG_EXPOSURE should be sufficient to define the amount of exposure the patient received, that is, the dose. Although OMOP researchers to date have calculated dose in an ad-hoc fashion for each study question, their research will greatly benefit in reproducibility, reliability, and transparency from a standardized approach. Therefore, we aimed to develop a standardized methodology for daily dose calculations in OMOP CDM data sources and to implement it in OHDSI standardized analytics.

2 | METHODS

We built a methodology to calculate daily dose of a drug depending on a pattern of used units in DRUG_STRENGTH in which the strength

of a drug product is stored as amount, mainly for non-divisible dose forms (e.g., pills, capsules, suppositories) or in numerator plus denominator, specifying a drug concentration, mainly used for divisible dose forms (e.g., liquids, aerosols). Thus, the same ingredient is stored in different ways in relation to the dose form of the drug product. Data S1 depicts and explains the organization of the DRUG_STRENGTH reference table in detail given a few drug product examples.

Only clinically relevant units were used in our calculations, others such as homeopathic dosages or bacteria units were disregarded. Clinically relevant units of the amount and numerator were standardized to international unit (IU), milliequivalent, milligram, and milliliter. Denominator units were standardized to hour, milligram, milliliter, actuation, and square centimeter. This selection resulted in 41 structured patterns, which we further call drug strength patterns. In DRUG_STRENGTH, a drug can only populate either the amount fields or both the numerator and denominator fields; other combinations are not possible. Table 1 depicts these options schematically in DRUG_STRENGTH with detailed information available for the 41 identified drug strength patterns available in Data S2.

To develop dose formulas, we assessed the OMOP drug concepts associated with the previously identified 41 drug strength patterns in the DRUG_EXPOSURE table and their corresponding DRUG_STRENGTH information in three databases: CPRD GOLD,⁷ IPCI,⁸ and P +⁹ (database information in Table 2 and Data S3). We carried out the assessment per administration route (oral, injection, inhalation, transdermal, transmucosal, and topical) because route may influence dose and to allow for the possibility of defining different dose formulas for different routes. This review of how drug information was stored and used for most common drug concepts led to the formation of dose formulas. Thus, the resulting dose formulas were grounded in the OHDSI Standardized Vocabularies and have been confirmed by clinical assessments. Subsequently, we investigated the proportion of recorded OMOP drug concepts in the DRUG_EXPOSURE table for which we would be able to calculate dose with our suggested dose formulas. Finally, we selected six select pharmaceutical ingredients to test our formulas' dose results against an external benchmark, the

TABLE 1 Available options from DRUG_STRENGTH for the 41 identified drug strength patterns (shown with respective pattern name and dose unit). Detailed information available in Supplementary material 2.

| Amount value ^a | Amount unit ^b | Numerator value ^c | Numerator unit ^d | Denominator value ^e | Denominator unit ^f | Pattern name | Dose unit |
|---------------------------|--------------------------|------------------------------|-----------------------------|--------------------------------|-------------------------------|-------------------------------|-----------|
| Numeric | Present | NA | NA | NA | NA | fixed_amount_unit | unit |
| NA | NA | Numeric | Present | Numeric | Present | numUnit_per_denomUnit | numUnit |
| NA | NA | Numeric | Present | NA | Present | numUnit_per_missing_denomUnit | numUnit |

Note: NA: missing value; numUnit and denomUnit represent the numerator and denominator unit, respectively; present is a placeholder for clinically relevant units depicted in Data S2.

^aAmount value: numeric value or missing information on the amount of active ingredient contained within the drug product.

^bAmount unit: The concept representing the unit of measure for the amount of active ingredient contained within the drug product.

^cNumerator value: The concentration of the active ingredient contained within the drug product.

^dNumerator unit: The concept representing the Unit of measure for the concentration of active ingredient.

^eDenominator value: The amount of total liquid (or other divisible product, such as ointment, gel, spray, etc.). If NA, the assumption is that denominator value equals one.

^fDenominator unit: The Concept representing the denominator unit for the concentration of active ingredient.

TABLE 2 Databases participating in this study and their contribution.

| Full name | Abbreviation | Data type | Location | Population | Contribution |
|---|--------------|--|------------------|--|--|
| Clinical Practice Research Datalink | CPRD GOLD | Primary care EHR | United Kingdom | Representative sample of 7% of inhabitants | Derivation of dose formulas (proof of concept) and calculating daily dose for six selected pharmaceutical ingredients (benchmarking) |
| Integrated Primary Care Information | IPCI | Primary care EHR | The Netherlands | Representative sample of 16% of inhabitants | |
| PharMetrics® Plus for Academics | P+ | Claims from primary care and secondary care | United States | 107 million insured patients not representative of the US population (higher proportion of insured patients aged 0–18 and a lower proportion of insured patients aged >34) | |
| University of Tartu dataset of health data | MAITT | Claims and EHR from primary and secondary care | Estonia | Representative sample of 10% of inhabitants | Calculating daily dose for six selected pharmaceutical ingredients (benchmarking only) |
| IQVIA Disease Analyzer Germany | IQVIA DA | Primary care and secondary care EHR | Germany | 42.5 million patients largely representative of all primary and secondary care outpatients in Germany | |
| IQVIA Longitudinal Patient Database Belgium | IQVIA LPD | Primary care EHR | Belgium | 1.1 million inhabitants overrepresenting youngest and oldest age groups | |
| Multicenter Integrated Hospital Information System Parc Salut | IMASIS | Secondary care EHR | Barcelona, Spain | Hospital based information on 1.6 million patients | |

Note: Detailed information on databases is available in Data S3.

Abbreviation: EHR, electronic health records.

World Health Organization's (WHO) Defined Daily Dose (DDD).¹⁰ The WHO DDD is defined as “the assumed average maintenance dose per day for a drug used for its main indication in adults.”¹¹ The chosen ingredients—metformin, enoxaparin, furosemide, salmeterol, tiotropium, and fentanyl—are diverse in units, administration route, and the setting in which they are used. Using the dose formula for the specific route and drug strength pattern of the drug concepts of interest, we calculated overall median daily dose and the 25th and 75th percentile, and additional stratified results by route and drug strength pattern. The benchmarking was carried out in the three databases from which we derived the dose formulas and additionally in four European databases representing a variety of healthcare settings: MAITT,¹² IQVIA DA,¹³ IQVIA LPD, and IMASIS¹⁴ (database information in Table 2 and Data S3). All databases had been previously mapped to the OMOP CDM. The selected databases have overall good data quality with sufficient information on the needed variables to proceed with dose calculations (i.e., quantity and duration).

2.1 | Study code

The code used to obtain the drug strength patterns, assess the dose formula's applicability, and daily dose benchmarking is freely

available in a public GitHub repository: <https://github.com/oxford-pharmacoepi/DailyDoseRouteValidationStudy>.

3 | RESULTS

The resulting dose formulas led to the categorization of the 41 identified drug strength patterns into three groups: fixed amount formulation patterns, time-based formulation patterns, and concentration formulation patterns. The dose formulas are depicted in Table 3. Another outcome from the clinical review was that the definition of the quantity field was overburdened with different usages. We observed that the use of quantity differed depending on drug strength pattern and dose form. The reason is that in the current CDM version, quantity is used for manifold purposes which shall be resolved in a future CDM version. Our dose formula stem from how the quantity field was being used as observed during our clinical review (some consistent with current definitions, some not). Thus, we state our definition of quantity together with each group.

The first group, fixed amount formulation patterns, contain those patterns that had amount_value numeric, amount_unit present, numerator_unit missing, and denominator_unit missing. These patterns comprise drugs where the drug strength is measured in a fixed

TABLE 3 Dose formulas for the three groups of drug strength patterns.

| Pattern group | DRUG_STRENGTH information | Daily dose formula | Usual dose forms |
|------------------------------------|---|--|---|
| Fixed amount formulation patterns | amount_value numeric, amount_unit present, missing numerator_unit, and missing denominator_unit | $\frac{\text{Amount value} \times \text{Quantity}}{(\text{Drug exposure end date} - \text{Drug exposure start date} + 1)^a}$ | Non-divisible dose forms, for example, pills, capsules, suppositories, and patches |
| Time-based formulation patterns | numerator_value numeric, numerator_unit present, and denominator_unit “hour” | Denominator value ≤ 24 h: Numerator value Denominator value > 24 h or missing denominator value (i.e., denominator value = 1): $\frac{\text{Numerator value} \times 24}{\text{Denominator value}}$ | Non-divisible dose forms particularly dosed by time, for example, patches and extended releases solid oral dose forms |
| Concentration formulation patterns | numerator_value numeric, numerator_unit present, and denominator_unit not “hour” | $\frac{\text{Numerator value} \times \text{Quantity}}{(\text{Drug exposure end date} - \text{Drug exposure start date} + 1)^b}$ | Divisible dose forms, for example, oral/inhalable solutions and injectables |

Note: Amount value, numerator value, and denominator value come from the DRUG_STRENGTH table. Quantity, drug exposure start date, and drug exposure end date come from the DRUG_EXPOSURE table. The dose formulas were implemented in the DrugUtilisation R package,¹⁵ which is freely available under the Apache License (Version 2.0) and can be obtained from CRAN (<https://cran.r-project.org/web/packages/DrugUtilisation/index.html>).

^aThe quantity is the number of units/tablets/capsules/others prescribed or dispensed.

^bOur clinical assessment suggested that the quantity depended on whether denominator value was missing (=1) or not. In cases where denominator value was missing, the quantity was mainly populated by giving the total volume/weight/others of the product prescribed or dispensed. In cases where the denominator value was not missing, we mainly saw single or multiple unit packages, and the quantity was populated with the number of bottles/units/sachets/others of the product prescribed or dispensed.

amount of a certain ingredient. This can be best thought of with regards to pills or suppositories for example but includes other dose forms as well. The amount_unit defines the unit of the calculated daily dose for all fixed amount formulation patterns. The quantity here is the number of units/tablets/capsules/others prescribed or dispensed.

The second group, time-based formulation patterns, contain those patterns that had numerator_value numeric, numerator_unit present, and denominator_unit “hour.” These patterns comprise drugs where the drug strength is measured by time. Thus, the patient receives an amount of a certain ingredient per time unit. The drugs found in this group are transdermal patches and extended-release dose forms for example. We suggest two different dose formulas depending on the presence of the denominator value and its numeric value. On one hand, if the denominator value is ≤ 24 h the daily dose is the numerator_value because there is most likely no repetition of the drug if given for less than 24 h (most likely the remaining hours until 24 h represent a break from the exposure). On the other hand, if the denominator value is > 24 h or missing, the drug concentration per hour is multiplied by 24 and divided by the denominator_value. For missing denominator_value, the assumption is that the denominator value equals 1. The numerator_unit defines the unit of the calculated daily dose for all time-based formulation patterns. The quantity here is not relevant because it is not part of the equation. The third group, concentration formulation patterns, contain those patterns that had numerator_value numeric, numerator_unit present, and denominator_unit any unit but “hour.” These patterns include drugs where the drug strength are given as a concentration. Thus, the patient receives a certain “amount” of ingredient per administered volume of a solution or per administered actuation of a device for example. The numerator_unit defines the unit of the calculated daily dose for all concentration formulation patterns. Our clinical assessment suggested

that the quantity depended on whether the denominator value was missing (=1) or not. In cases where the denominator value was missing, the quantity was mainly populated by giving the total volume/weight/others of the product prescribed or dispensed. An example is quantity 100 for the drug_concept_id: 1713520 “amoxicillin 25 mg/mL oral suspension”, that is, 100 mL. In cases where the denominator value was not missing, we mainly saw single or multiple unit packages, and the quantity was populated with the number of bottles/units/sachets/others of the product prescribed or dispensed. An example is quantity 15 for the drug_concept_id: 40708507 “1000 mg Estriol 0.001 mg/mg topical cream”, that is, 15 units.

The dose formulas were implemented in the DrugUtilisation R package,¹⁵ which is freely available under the Apache License (Version 2.0) and can be obtained from CRAN (<https://cran.r-project.org/web/packages/DrugUtilisation/index.html>). The package includes documentation on how to calculate and summarize dose for a drug cohort using the addDailyDose() and addDrugUse() function,¹⁶ respectively. Furthermore, the package allows flexible handling of gaps and overlapping exposure periods to be decided by the user.

The applicability of the drug formulas among the drug concepts in the DRUG_EXPOSURE table was high and ranged from 85.8% in IQVIA DA Germany to 96.9% in MAITT. IMASIS with only 52.1% was the only outlier due to a large amount of drug concepts in the amount formulation patterns without a numeric amount value and hence invalid for dose calculation.

Overall results of calculated median daily dose of metformin, enoxaparin, furosemide, salmeterol, tiotropium, and fentanyl are depicted in Table 4. We observed that most databases yielded median daily doses similar to the WHO DDD for all ingredients (within a range of a factor of 3). Outliers were observed for P+ for enoxaparin, for CPRD GOLD and P+ for salmeterol, for IQVIA LPD and IMASIS

TABLE 4 Overall daily dose calculations per ingredient per database.

| Ingredient | Median daily dose (IQR) | | | | | |
|--------------|----------------------------|--|---------------------------|----------------------------|---|--|
| | Metformin | Enoxaparin | Furosemide | Salmeterol | Tiotropium | Fentanyl |
| WHO DDD | 2 g oral | 2000 IU injection (100 IU = 1 mg) | 40 mg oral / injection | 0.1 mg inhalation | 0.01 mg inhalable powder, 0.005 mg inhalable solution | 0.6 mg nasal / sublingual, 1.2 mg transdermal |
| CPRD GOLD | 1700 mg (1000– 2000) | 60 mg (40–120) | 40 mg (20– 40) | 0.00 mg (0.00– 0.10) | 0.018 mg (0.018– 0.019) | 0.6 mg (0.6–1.2) |
| IPCI | 1000 mg (938–2000) | 40 mg (22–100) , 4000 IU (1521–4000) | 40 mg (20– 40) | 0.10 mg (0.10– 0.10) | 0.018 mg (0.005– 0.018) | 1.4 mg (0.7–2.8) |
| MAITT | 1700 mg (1000– 2000) | 2667 IU (1333–4000) | 60 mg (40– 60) | 0.20 mg (0.10– 0.20) | 0.010 mg (0.005– 0.010) | 0.6 mg (0.6–1.2) |
| P+ | 1000 mg (500–2000) | 160 mg (64–160) | 40 mg (20– 40) | 6.00 mg (6.00– 6.00) | 0.018 mg (0.018– 0.020) | 0.1 mg (0.1–0.2) |
| IQVIA DA | 2000 mg (1500– 2000) | 27 mg (20–40) , 2.7 IU (1.3–4) | 40 mg (40– 40) | 0.15 mg (0.15– 0.15) | 0.018 mg (0.015– 0.054) | 0.3 mg (0.1–0.6) |
| IQVIA LPD | 1000 mg (850–1700) | 8000 IU (4000–16'000) | 40 mg (40– 40) | 0.04 mg (0.04– 0.17) | 0.54 mg (0.018– 0.54) | NA |
| IMASIS | 850 mg (425– 850) | 60 mg (20–60) | 40 mg (40– 40) | 0.04 mg (0.04– 3.00) | 0.54 mg (0.54– 0.054) | 5 mg (0.6–7.5) |

Note: Bold values indicate the results of the larger strata.

Abbreviations: CPRD GOLD, Clinical Practice Research Datalink GOLD; DA, Disease Analyzer; IMASIS: Multicenter Integrated Hospital Information System; IPCI, Integrated Primary Care Information; LPD, Longitudinal Patient; MAITT, University of Tartu dataset of health data; NA, not available; P+, PharMetrics® Plus for Academics.

for tiotropium, and for P+ and IMASIS for fentanyl. The overall daily dose calculations for tiotropium and fentanyl may not be informative because the WHO DDD is different for different dose forms and administration routes, respectively. Daily dose calculations stratified by route and additional by drug strength pattern for tiotropium are depicted in Table 5. The fixed amount formulation patterns relate to the inhalable powder capsule with a WHO DDD of 0.01 mg and the concentration formulation patterns relate to the inhalable solution with a WHO DDD of 0.005 mg. Most databases depicted a clear distinction of calculated daily dose per drug strength pattern for the two different dose forms. Moreover, the overall calculations for IQVIA LPD (which happened to be an outlier) were masking the less frequent inhalable solution for which the drug strength pattern level calculation yielded a value identical to that of the WHO DDD. Daily dose calculations stratified by route and additional by drug strength pattern for fentanyl are depicted in Table 6. IPCI and MAITT yielded median daily doses close to the WHO DDD for all administration routes. All databases had at least one route and pattern strata that yielded correct WHO DDD.

All results of calculated median daily dose for all ingredients including those for strata of route and strata of route and drug

strength pattern are shown in the Data S4. For the majority of cases, the large and therefore important strata yielded correct daily dose calculations. However, we observed quite some variation of median daily doses between administration routes (which may be expected for some ingredients) and even larger variation between different drug strength patterns of the same route (which is unexpected). Yet, the distribution of the proportion of drug records between drug strength patterns was diverse and some were smaller than 1% and are therefore negligible.

4 | DISCUSSION

We developed and implemented a standardized way to calculate drug daily dose in OMOP CDM data sources though utilizing the DRUG_STRENGTH reference table and quantity and duration from DRUG_EXPOSURE. Our approach allows overall daily dose calculation and stratification by route, and additionally by drug strength information. This approach makes the dose calculation process transparent, reliable, and reproducible. Furthermore, the daily dose calculation implementation is compatible with the existing OHDSI analytical pipeline.

TABLE 5 Daily dose calculations of tiotropium (WHO DDD: 0.01 mg inhalable powder / 0.005 mg inhalable solution) stratified by route, and by pattern and route.

| | Stratification by route | | | Additional stratification by pattern | | |
|-----------|-------------------------|------|-------------------------|--------------------------------------|------|-------------------------|
| | Stratum | % | Daily dose median (IQR) | Stratum | % | Daily dose median (IQR) |
| CPRD GOLD | inhalation | 100 | 0.018 mg (0.018–0.019) | fixed_amount_mg | 86.1 | 0.018 mg (0.018–0.019) |
| | | | | mg_per_missing_actuation | 13.9 | 0.005 mg (0.005–0.005) |
| IPCI | inhalation | 100 | 0.018 mg (0.005–0.018) | fixed_amount_mg | 60.7 | 0.018 mg (0.018–0.018) |
| | | | | mg_per_actuation | 39.3 | 0.005 mg (0.005–0.005) |
| | | | | mg_per_missing_actuation | 0.0 | 0.000 mg (0.000–0.003) |
| | NA | 0.0 | | NA | 0.0 | |
| MAITT | inhalation | 100 | 0.010 mg (0.005–0.010) | mg_per_missing_actuation | 100 | 0.010 mg (0.005–0.010) |
| P+ | inhalation | 100 | 0.018 mg (0.018–0.020) | fixed_amount_mg | 51.7 | 0.018 mg (0.018–0.018) |
| | | | | mg_per_actuation | 48.3 | 0.020 mg (0.020–0.020) |
| IQVIA DA | inhalation | 87.0 | 0.018 mg (0.015–0.054) | fixed_amount_mg | 58.4 | 0.036 mg (0.018–0.054) |
| | | | | mg_per_actuation | 20.7 | 0.010 mg (0.005–0.015) |
| | | | | mg_per_ml | 7.8 | 0.000 mg (0.000–0.000) |
| | NA | 13.0 | | NA | 13.0 | |
| IQVIA LPD | inhalation | 100 | 0.540 mg (0.036–0.540) | fixed_amount_mg | 69.2 | 0.540 mg (0.036–0.540) |
| | | | | mg_per_actuation | 30.8 | 0.005 mg (0.003–0.45) |
| IMASIS | inhalation | 31.8 | 0.540 mg (0.540–0.054) | mg_per_actuation | 31.8 | 0.540 mg (0.540–0.054) |
| | NA | 68.3 | | NA | 68.3 | |

Note: The amount or numerator unit of the pattern defines the unit of the calculated daily dose.

Abbreviations: CPRD GOLD, Clinical Practice Research Datalink GOLD; DA, Disease Analyzer; IMASIS, Multicenter Integrated Hospital Information System; IPCI, Integrated Primary Care Information; LPD, Longitudinal Patient; MAITT, University of Tartu dataset of health data; NA, no calculations or no clinically relevant unit; P+, PharMetrics® Plus for Academics.

Obtained median daily dose results were largely consistent with WHO DDD. A study comparing dose calculations among US data with WHO DDD yielded 86% of prescriptions within a factor three of the WHO DDD, yet they did not disclose their dose formulas.¹⁷ The cases for which the daily dose calculation did not match the expected results in our assessment may shed light on data mapping problems, inconsistencies in the source data, or prescription patterns truly deviating from the recommendations. One obvious source of error might be the mapping of the source drug code to OMOP's standard concepts, which are the basis for DRUG_STRENGTH. These mappings may be insufficiently granular (e.g., mapping to concepts that carry only ingredient but no strength information) for example. Such a mapping problem can be detected by running the DARWIN EU developed DrugExposureDiagnostic¹⁸ R package which assesses the DRUG_EXPOSURE table and with which the mapping level can be quickly checked. Other mapping problems can be overcome at the mapping stage with knowledge of the data, the country, and the healthcare system (e.g., if there is only one strength on the market). One could claim that the strength of drug products is a known fact, but since there is no general global compendium available except the DRUG_STRENGTH table, inaccuracies in the latter will result in incorrect results. Moreover, our dose formulas rely on correct information in the start and end date of drugs (i.e., duration) and the quantity prescribed/dispensed. These values can also be checked using the DrugExposureDiagnostic R package. And we strongly recommend doing so to obtain sensible daily dose estimates. Finally, some

inconsistencies between our daily dose calculation and the WHO DDD can be explained by the fact that the DDD, as an average value, does not necessarily correspond to the recommended or prescribed daily dose and that our results are therefore true deviations depending on the healthcare setting and country. Yet, the WHO DDD is the best available benchmark.

Dose calculations were conducted overall and additionally stratified by route and drug strength patterns. This detailed stratification helps to locate the drug concepts that are behind the respective dose calculations. Depending on what the dosing information is used for, for example, if different doses are compared, the stratification allows for restricting the drug strength patterns or even drug concepts to those providing reliable calculations, that is, to those for which exists high certainty that the source is correct and that they have been correctly mapped. Moreover, this detailed investigation into drug daily dose is also a basis for valuable feedback for all databases to question their data quality. A fixable problem that we have observed during the process are differences in the population of the quantity field in the DRUG_EXPOSURE table, which is present in most dose formulas. To that end, daily dose calculations within OMOP CDM can also be seen as a quality indicator for a database because each data component needs to be at the right place upstream before data can produce correct daily dose.

The current formulas' ability to apply to over 85% of recorded drug concepts in most databases is promising and covers likely all relevant drugs. Yet, high applicability will not automatically lead to

TABLE 6 Daily dose calculations of fentanyl (WHO DDD: 0.6 mg nasal/sublingual, 1.2 mg transdermal) stratified by route and unit, and by pattern and route. The amount or numerator unit of the pattern defines the unit of the calculated daily dose.

| | Stratification by route | | | Additional stratification by pattern | | |
|-----------|-------------------------|------------------|-------------------------|--------------------------------------|------------------|-------------------------|
| | Stratum | % | Daily dose median (IQR) | Stratum | % | Daily dose median (IQR) |
| CPRD GOLD | transdermal | 96.9 | 0.6 mg (0.6–1.2) | mg_per_missing_h | 96.9 | 0.6 mg (0.6–1.2) |
| | buc./subl. | 3.0 | 3.0 mg (0.8–12.0) | fixed_amount_mg | 3.0 | 3.0 mg (0.8–12.0) |
| | nasal | 0.0 | 1.0 mg (0.5–3.2) | mg_per_missing_actuation | 0.0 | 1.0 mg (0.5–3.2) |
| | injection | 0.0 | 2.0 mg (1.0–5.0) | mg_per_ml | 0.0 | 2.0 mg (1.0–5.0) |
| | inhalation | 0.0 | 2.0 mg (0.8–3.2) | mg_per_missing_actuation | 0.0 | 2.0 mg (0.8–3.2) |
| | NA | 0.0 | | NA | 0.0 | |
| IPCI | transdermal | 88.7 | 1.4 mg (0.9–2.8) | mg_fixed_amount | 88.7 | 1.4 mg (0.9–2.8) |
| | buc./subl. | 4.7 | 0.6 mg (0.3–1) | mg_fixed_amount | 4.7 | 0.6 mg (0.3–1) |
| | | | | mg_per_ml | 0.0 | 0.0 mg (0.0–0.1) |
| | nasal | 4.3 | 0.4 mg (0.1–0.6) | mg_per_actuation | 4.2 | 0.4 mg (0.1–0.6) |
| | | | | mg_fixed_amount | 0.1 | 0.1 mg (0.1–0.3) |
| | | | | mg_per_missing_actuation | 0.0 | 0.0 mg (0.0–0.0) |
| | | | mg_fixed_amount | 0.3 | 0.0 mg (0.0–0.1) | |
| | 1.9 | | NA | 1.9 | | |
| MAITT | transdermal | 72.5 | 0.6 mg (0.6–1.2) | mg_per_missing_h | 72.5 | 0.6 mg (0.6–1.2) |
| | buc./subl. | 27.3 | 0.2 mg (0.1–0.3) | mg_per_missing_actuation | 27.3 | 0.2 mg (0.1–0.3) |
| | nasal | ~ | | fixed_amount_mg | ~ | |
| P+ | injection | 28.7 | 0.1 mg (0.1–0.1) | mg_per_ml | 28.4 | 0.1 mg (0.1–0.1) |
| | | | | mg_per_missing_ml | 0.3 | 0.1 mg (0.1–0.2) |
| | transdermal | 6.9 | 1.2 mg (0.6–1.8) | fixed_amount_mg | 6.9 | 1.2 mg (0.6–1.8) |
| | buc./subl. | 0.1 | 3.0 mg (1.6–6.4) | fixed_amount_mg | 0.1 | 3.2 mg (1.6–6.7) |
| | | | | mg_per_missing_actuation | 0.0 | 2.3 mg (1.2–3) |
| | inhalation | 0.0 | 0.2 mg (0.1–0.2) | mg_per_missing_actuation | 0.0 | 0.2 mg (0.1–0.2) |
| | 0.1 | | NA | 0.1 | | |
| IQVIA DA | transdermal | 96.2 | 0.3 mg (0.1–0.6) | mg_per_h | 51.0 | 0.0 mg (0.0–0.1) |
| | | | | mg_per_missing_h | 46.0 | 0.6 mg (0.3–1.2) |
| | | | | fixed_amount_mg | 0.0 | 0.0 mg (0.0–0.0) |
| | buc./subl. | 3.4 | 0.3 mg (0.2–0.9) | fixed_amount_mg | 3.4 | 0.3 mg (0.2–0.9) |
| | nasal | 0.4 | 0.1 mg (0.0–0.2) | mg_per_ml | 0.3 | 0.0 mg (0.0–0.2) |
| | | | | mg_per_missing_actuation | 0.0 | 0.2 mg (0.2–0.2) |
| | | | | mg_per_missing_ml | 0.0 | 0.0 mg (0.0–0.0) |
| | injection | 0.1 | 0.0 mg (0.0–0.2) | mg_per_ml | 0.1 | 0.0 mg (0.0–0.2) |
| oral | 0.0 | 1.5 mg (0.3–1.5) | fixed_amount_mg | 0.0 | 1.5 mg (0.3–1.5) | |
| | 0.0 | | NA | 0.0 | | |
| IQVIA LPD | transdermal | 99.7 | NA | mg_per_missing_h | 99.7 | NA |
| | | | | fixed_amount_mg | 0.1 | NA |
| | subl./buc. | 0.1 | 0.1 mg (0.1–0.1) | fixed_amount_mg | 0.0 | 0.1 mg (0.1–0.1) |
| | 0.3 | | NA | 0.3 | | |
| IMASIS | injection | 68.9 | 5.0 mg (2.5–7.5) | mg_per_missing_ml | 68.9 | 5 mg (2.5–7.5) |
| | transdermal | 21.3 | 0.6 mg (0.3–1.2) | mg_per_h | 18.2 | 0.6 mg (0.3–1.2) |
| | | | | mg_per_missing_h | 3.1 | 0.6 mg (0.3–2.4) |
| | subl./buc. | 1.7 | 0.1 mg (0.0–0.1) | fixed_amount_mg | 1.7 | 0.1 mg (0.0–0.1) |
| | NA | 7.7 | | NA | 7.7 | |

Note: ~ Results with less than 5 records are suppressed.

Abbreviations: CPRD GOLD, Clinical Practice Research Datalink GOLD; DA, Disease Analyzer; IMASIS, Multicenter Integrated Hospital Information System; IPCI, Integrated Primary Care Information; LPD, Longitudinal Patient; MAITT, University of Tartu dataset of health data; NA, no calculations or no clinically relevant unit; P+, PharMetrics® Plus for Academics.

meaningful daily dose calculations, because from a pharmacological point of view, a sensible drug daily dose calculation requires regular and continuous drug use by the patient. Yet, we do not know if the patient collected and used the drug as prescribed. Moreover, our daily dose calculation only covered the intended dose by the prescriber and what the patient ingests is impossible to know. Furthermore, daily dose calculations for medications that are used on demand would not yield a clinically meaningful daily exposure because it would give an average of the exposure rather than the actual daily intermittent exposure. Thus, an approach to quantify drug dose of intermittent medication use can be through cumulative dose calculations for example.

The major strength of this study is its systematic and structured approach toward calculation of daily dose within the OMOP CDM. The creation of structured drug strength patterns, assessment of OMOP drug concepts in DRUG_EXPOSURE, and comprehensive benchmarking against WHO DDD standards enhance the reliability and transparency of the methodology, and finally the reproducibility of the calculated results. Thus, our dose formula and its implementation in the DrugUtilisation R package allow dose calculation within OMOP CDM on a global scale (given the diagnostic assessment suggests dose calculations for the desired study question). Moreover, possible stratification by route and drug strength pattern (often aligns with formulation) helps with country/market and product specific interpretation of results.

However, it is essential to acknowledge certain limitations. If drug concepts of the same ingredient have different units (for example enoxaparin, IU and mg), then the dose calculations yield a result per unit with the current implementation. A general harmonization of units is not possible here because the conversion of IU to mg depends on the ingredient. Furthermore, while the applicability of dose formulas is high in most databases, there was one database with below 50%, mainly due to missing numeric values in the amount formulation patterns. The reason is that this database has different drug types—according to the CDM vocabulary⁶—available in their system with different levels of granularity. This is something that would be detected in the diagnostics step and actions can be taken to only use the useful drug type when conducting the study. Additionally, this study's benchmarking step may not encompass the entire spectrum of drugs in clinical practice although we tried our best to select ingredients as varied as possible. The same is true for the participating databases, while we tried to be comprehensive in the database selection, it may not be generalizable to all databases in the OMOP CDM. Yet, the use of the OHDSI Standardized Vocabularies should ascertain use of our dose formulas across the entire OMOP CDM world. Furthermore, there will be circumstances when using the “signetur” (sig field in DRUG_EXPOSURE) may yield more precise dose calculations. However, this lacks a standard representation in CDM do date. Therefore, we did not use this information in the dose formulas but plan to amend the dose formula for different circumstances with time. Finally, OMOP has guidelines¹⁹ on how the data has to be mapped and what a successful mapping means assessed through the data quality dashboard.²⁰ However, given different data types and sources, not all OMOP databases

will have accurate information to calculate dose (especially with regards to exposure duration). Therefore, the diagnostic assessment of the data is utterly important before dose calculation.

To conclude, we provided a standardized methodology for calculating daily drug doses within the OMOP CDM that was benchmarked against existing universal measures of drug consumption. The suggested dose formulas determined through rigorous evaluation enhance the reliability, transparency, and reproducibility of daily doses in pharmacoepidemiologic studies.

AUTHOR CONTRIBUTIONS

TB and MC conceptualised this study. EB, KLG, and MC authored the DrugUtilisation software package. MC, KLG, and TB wrote the R script for this study. MC, TB, MR, MM, JG, SS, DV, MAM, JMR, AL, MO, and RK utilized the script against their data and were therefore responsible for performing this study. AG, AMJ, CR, KB, LB, and TB provided epidemiological and clinical expertise. TB wrote the initial version of the manuscript. All authors interpreted the results, critically reviewed the manuscript and approved the final version for submission and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

ACKNOWLEDGEMENTS

We acknowledge the contributions of Albert Prats Uribe in discussing clinically relevant units and first dose formulas. Furthermore, we acknowledge Romain Griffier, Talita Duarte Salles, and Núria Mercadé Besora for contributing to and commenting on an earlier version of this work.

FUNDING INFORMATION

Development of the dose finding process was partially funded by the European Medicines Agency as part of the Data Analysis and Real-World Interrogation Network (DARWIN EU). This manuscript represents the views of the DARWIN EU Coordination Centre only and cannot be interpreted as reflecting the views of the European Medicines Agency or the European Medicines Regulatory Network. Further partial funding comes from the European Health Data and Evidence Network (EHDEN). The study funders had no role in the data collection, conceptualisation, design, analysis, interpretation, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST STATEMENT

DPA's department has received grant/s from Amgen, Chiesi-Taylor, Lilly, Janssen, Novartis, and UCB Biopharma. His research group has received consultancy fees from AstraZeneca and UCB Biopharma. Amgen, Astellas, Janssen, Synapse Management Partners and UCB Biopharma have funded or supported training programs organized by DPA's department. MR and MM work for a research group that—in the past 5 years—received unconditional research grants from, GSK, Amgen, Chiesi, AstraZeneca, UCB, J&J, the European Medicines Agency and the Innovative Medicines Initiative. DV, KB, and JG are employed by IQVIA.

DATA AVAILABILITY STATEMENT

In accordance with current European and national law, the data used in this study is only available for the researchers participating in this study. Thus, we are not allowed to distribute or make publicly available the data to other parties. The source code of the R script is publicly available at: <https://github.com/oxford-pharmacoepi/DailyDoseRouteValidationStudy>. The source code of the DUS R package in which the dose calculation was implemented is publicly available at: <https://github.com/darwin-eu/DrugUtilisation>.

ETHICS STATEMENT

The protocol for this research was approved by CPRD's Research Data Governance (RDG) Process (protocol number 22_002354), the IPCI review board (protocol number 2019–13), the Hospital del Mar Barcelona Ethics Committee (CEIm 2022/10693), and the Estonian committee on bioethics and human research (approval 1.1–12/653).

ORCID

Theresa Burkard  <https://orcid.org/0000-0003-1313-4473>

Daniel Prieto-Alhambra  <https://orcid.org/0000-0002-3950-6346>

REFERENCES

- Reich C, Ostropolets A, Ryan P, et al. OHDSI standardized vocabularies—a large-scale centralized reference ontology for international data harmonization. *J Am Med Inform Assoc*. 2024;31:583–590. doi:10.1093/jamia/ocad247
- Voss EA, Makadia R, Matcho A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*. 2015;22(3):553–564. doi:10.1093/jamia/ocu023
- Hripcsak G, Duke JD, Shah NH, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574–578. <http://www.ncbi.nlm.nih.gov/pubmed/26262116><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4815923>
- Darwin EU. coordination centre. DARWIN EU 2023. Accessed October 30. <https://www.darwin-eu.org/>
- OHDSI. OMOP CDM Drug Strength Table. Accessed October 30. https://ohdsi.github.io/CommonDataModel/cdm54.html#DRUG_STRENGTH. 2023.
- OHDSI. OMOP CDM Drug Exposure Table. Accessed October 30. https://ohdsi.github.io/CommonDataModel/cdm54.html#DRUG_EXPOSURE. 2023.
- Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827–836. doi:10.1093/ije/dyv098
- De Ridder MAJ, De Wilde M, De Ben C, et al. Data resource profile: the integrated primary care information (IPCI) database The Netherlands. *Int J Epidemiol*. 2022;51(6):E314–E323. doi:10.1093/ije/dyac026
- IQVIA. IQVIA PharMetrics® Plus for Academics. 2022.
- World Health Organisation. ATC/DDD Index. 2023 Accessed October 30 https://www.whooc.no/atc_ddd_index/
- World Health Organization (WHO). Defined Daily Dose (DDD). Accessed April 5, <https://www.who.int/tools/atc-ddd-toolkit/about-ddd> 2024.
- Oja M, Tamm S, Mooses K, et al. Transforming Estonian health data to the observational medical outcomes partnership (OMOP) common data model: lessons learned. *JAMIA Open*. 2023;6(4):1–10. doi:10.1093/jamiaopen/oad100
- 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;56(10):459–466. doi:10.5414/CP203320
- Mayer MA, Furlong LI, Torre P, et al. Reuse of EHRs to support clinical research in a hospital of reference. *Stud. Health Technol. Inform*. 2015;210. IOS Press:224–226. doi:10.3233/978-1-61499-512-8-224
- Catala M, Mike D, Guo Y, Lopez-Guell K. DrugUtilisation R Package. Accessed April 5 <https://darwin-eu-dev.github.io/DrugUtilisation/index.html> 2024.
- Vignette addDailyDose() and addDrugUse(). Accessed April 5, https://darwin-eu-dev.github.io/DrugUtilisation/articles/a04_addDrugInfo.html. 2024.
- Bodenreider O, Rodriguez LM. Analyzing U.S. prescription lists with RxNorm and the ATC/DDD index. *AMIA Annu Symp Proc*. 2014; 297–306.
- Inberg G, Burn E, Guo Y, et al. DrugExposureDiagnostics. Accessed April 19, <https://darwin-eu-dev.github.io/DrugExposureDiagnostics/> 2024.
- OHDSI. OMOP CDM. Accessed April 12 <https://ohdsi.github.io/CommonDataModel/index.html> 2024.
- Sadowski K, Blacketer C, Londhe A, et al. Data Quality Dashboard. Accessed April 12, <https://ohdsi.github.io/DataQualityDashboard/> 2024.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Burkard T, López-Güell K, Gorbachev A, et al. Calculating daily dose in the Observational Medical Outcomes Partnership Common Data Model. *Pharmacoepidemiol Drug Saf*. 2024;33(6):e5809. doi:10.1002/pds.5809