

ORIGINAL ARTICLE

Real-world reproducibility study characterizing patients newly diagnosed with multiple myeloma using Clinical Practice Research Datalink, a UK-based electronic health records database

Anouchka Seesaghur¹ | Natalia Petruski-Ivleva² | Victoria Banks^{1,3} |
Jocelyn Ruoyi Wang² | Pattra Mattox² | Edwin Hoeben¹ | Joe Maskell¹ |
David Neasham¹ | Shannon L. Reynolds² | George Kafatos¹

¹Department Center for Observational Research, Amgen Ltd, Uxbridge, UK

²Department Science, Aetion, Inc, Boston, Massachusetts

³VLB Contractors Ltd, Kent, UK

Correspondence

Anouchka Seesaghur, Department Center for Observational Research, Amgen Ltd, 1 Uxbridge Business Park, Sanderson Road, Uxbridge, Middlesex UB8 1DH, UK.
Email: anouchka@amgen.com

Funding information

Amgen

Abstract

Purpose: We evaluated the reproducibility of a study characterizing newly-diagnosed multiple myeloma (MM) patients within an electronic health records (EHR) database using different analytic tools.

Methods: We reproduced the findings of a descriptive cohort study using an iterative two-phase approach. In Phase I, a common protocol and statistical analysis plan (SAP) were implemented by independent investigators using the Aetion Evidence Platform® (AEP), a rapid-cycle analytics tool, and SAS statistical software as a gold standard for statistical analyses. Using the UK Clinical Practice Research Datalink (CPRD) dataset, the study included patients newly diagnosed with MM within primary care setting and assessed baseline demographics, conditions, drug exposure, and laboratory procedures. Phase II incorporated analysis revisions based on our initial comparison of the Phase I findings. Reproducibility of findings was evaluate by calculating the match rate and absolute difference in prevalence between the SAS and AEP study results.

Results: Phase I yielded slightly discrepant results, prompting amendments to SAP to add more clarity to operational decisions. After detailed specification of data and operational choices, exact concordance was achieved for the number of eligible patients (N = 2646), demographics, comorbidities (i.e., osteopenia, osteoporosis, cardiovascular disease [CVD], and hypertension), bone pain, skeletal-related events, drug exposure, and laboratory investigations in the Phase II analyses.

Conclusions: In this reproducibility study, a rapid-cycle analytics tool and traditional statistical software achieved near-exact findings after detailed specification of data and operational choices. Transparency and communication of the study design, operational and analytical choices between independent investigators were critical to achieve this reproducibility.

This work was presented at the 35th International Conference on Pharmacoepidemiology & Therapeutic Risk Management in Philadelphia, PA, August 24-28, 2019.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Pharmacoepidemiology and Drug Safety* published by John Wiley & Sons Ltd.

KEYWORDS

epidemiologic methods, healthcare database, observational study, replication, reproducibility, transparency

1 | INTRODUCTION

Regulatory, payer, and clinical decision-makers are increasingly adopting real-world evidence (RWE) derived from existing real-world data (RWD), such as longitudinal electronic health records (EHR) and administrative claims data, to support healthcare decisions.^{1,2} The European Medicine Agency (EMA) defines RWD as “*routinely collected data relating to a patient's health status or the delivery of health care from a variety of sources other than traditional clinical trials.*”¹ An important quality and acceptability criterion for RWE is the ability to replicate studies accurately. The latter requires full transparency of data processing, design and analytic choices, beyond what is typically included in publications. Increasingly, efforts are being made to improve the reproducibility of research by promoting transparency.³⁻⁷ A reproducible study, defined as “*independent investigators implementing the same methods in the same data and are able to obtain the same results (direct replication),*”⁴ requires complete access to the study data (i.e., analytic data sets) and methods for sharing codes and software environment as well as ensuring sufficiently detailed study documentation.⁷⁻⁹ It has also been argued that confirming the findings from observational studies bolsters the overall confidence of scientific evidence.^{7,10}

One of the most commonly used UK-based RWD sources is the Clinical Practice Research Datalink (CPRD) General Practice (GP) Online Data (GOLD).¹¹ In a validation study, the CPRD database demonstrated a high positive predictive value of various diagnoses and similar comparisons of incidence with other UK data sources.¹² The adoption of RWD sources including EHR provides an opportunity to generate clinical evidence in oncology^{13,14} and the CPRD database^{14,15} has been used to evaluate patients with multiple myeloma (MM), a hematological cancer of the bone marrow.¹⁶ Multiple myeloma, estimated to cause approximately 5700 incident cases of myeloma a year in the UK, is the second most common hematological malignancy in Europe and recognizing MM with nonspecific, multi-site symptoms is challenging. While several publications using CPRD exist for this patient population,¹⁷⁻¹⁹ the reproducibility of such work using CPRD has not been evaluated. Thus, we sought to evaluate the reproducibility of a study characterizing patients newly diagnosed with MM in CPRD using two different analytic tools.

2 | METHODS

An iterative two-phase approach was used to evaluate the reproducibility of a descriptive cohort study using primary care data. In Phase I, two teams of independent investigators implemented a common study protocol and statistical analysis plan (SAP) independently and in-parallel using different analytic tools, a cloud-based rapid-cycle

KEY POINTS

- Adoption of RWE derived from large databases has resulted in efforts to improve the reproducibility and transparency of research.
- Findings from our reproducibility study using common study documents but two different analytic tools highlighted that at a minimum adherence to standard protocol and reporting guidelines is necessary; however, complete transparency of the study design and operational decisions shared by independent investigators was also critical to a successful reproduction.
- Data source parameters and implementation decisions, especially the nuances of using common data models and enrollment and assessment windows, should be explicitly stated during study planning and protocol development.

analytic tool, (Aetion Evidence Platform® [AEP], version 3.12), and traditional, line-programming statistical software (SAS Enterprise Guide version 7.1). Phase II was an iteration of the analyses following the review and comparison of the Phase I findings, implementation decisions, and revisions to study documents (e.g., SAP).

2.1 | Data source

The population-based cohort study used the EHRs from the CPRD GOLD database, which contains anonymized longitudinal patient records from more than 600 UK-based and has primary care medical records of over 18 million patients.²⁰ CPRD contains demographic data, medical diagnoses, procedures (including laboratory investigations and results), and death information collected using a standardized form²¹; conditions, interventions and diagnostics use Read codes and medication prescribing recorded using the British National Formulary (BNF). We used CPRD data collected between January 1, 2004 and December 31, 2017 for this study.

Raw CPRD data and raw data converted to Observational Medical Outcomes Partnership (OMOP) CDM (version 4.0) were used for the AEP and SAS analyses, respectively. The latter includes only patients with CPRD data of acceptable quality for research and contains a total of approximately 15 million patients. The raw CPRD data contain approximately 18 million patients, regardless of the data quality.

The research protocol was reviewed and approved by the Independent Scientific Advisory Committee (ISAC, protocol 18_292).

2.2 | Study population

During the study period (January 1, 2004 to December 31, 2017), we identified patients newly diagnosed with MM between January 1, 2006 through December 31, 2016 to allow for a minimum of 2-year baseline and 1-year follow-up period. Follow-up began the day after cohort entry and ended at disenrollment from the GP practice, the last data collection date of the practice, death, or end of study period (Figure 1). The study cohort included patients ≥ 18 years who were registered at the GP for at least 2 years prior to diagnosis and had no history of solid tumors.

2.3 | Covariates

Several categories of covariates were assessed, including demographics (i.e., age and gender), comorbidities, baseline clinical conditions (i.e., bone pain, and skeletal-related events [SREs]), prescribed medications, laboratory investigations and laboratory results. Baseline demographics and comorbidities, including cardiovascular disease (CVD), chronic kidney disease (CKD), gout, hypertension, osteopenia, osteoporosis, osteoarthritis, and rheumatoid arthritis (RA), were assessed in all available data prior to MM diagnosis. Baseline clinical conditions, medication prescribed, and laboratory investigation were assessed in the 2 years prior to MM diagnosis. Baseline medications included were prescribed medications related to bone health and pain management, that is, denosumab, prolia, xgeva, bisphosphonates, and analgesics. Laboratory investigation included select diagnostic laboratory tests (calcium, serum creatinine, hemoglobin).

Refer to Data S1 for a list of variables, diagnosis definitions, and covariate definitions.

2.4 | Statistical analysis

We used descriptive statistics to report the number and proportion of patients meeting the pre-specified criteria and for binary and categorical covariates, and reported the mean, SD, and range for continuous covariates. The presence of bone complications, such as bone pain and SREs, were assessed during both baseline and in follow-up.

The two analytic approaches used were AEP, a rapid-cycle analytic tool that has been previously validated,^{3,22} and SAS software as the gold standard for statistical analyses. The reproducibility of study results was evaluated by calculating the match rate within broad categories (e.g., demographics, conditions, drug exposure, and procedures) and the absolute difference in prevalence between line-programming and rapid analytics for each individual characteristic included within the broad category. Match rate was defined as a percentage (rounded to the nearest tenth), calculated as the number of individual characteristics with an exact match in the estimated prevalence between the AEP and SAS results, divided by the total number of characteristics within each broad category. An absolute difference in prevalence of 0% and a match rate of 100% represented exact concordance. After the Phase I analysis, the two teams of independent investigators performed a careful review of the results to identify potential reasons for discrepancies and resolve them for the Phase II analysis. Thus, recommendations to promote transparency in study protocols and SAPs were developed.

3 | RESULTS

3.1 | Phase I: Initial analysis

In Phase I, the two analytic approaches yielded near-identical number of eligible patients with newly-diagnosed MM during the study period (N = 2646 in AEP vs. N = 2648 in SAS). Exact concordance for the distribution of gender and median age was observed; however, there

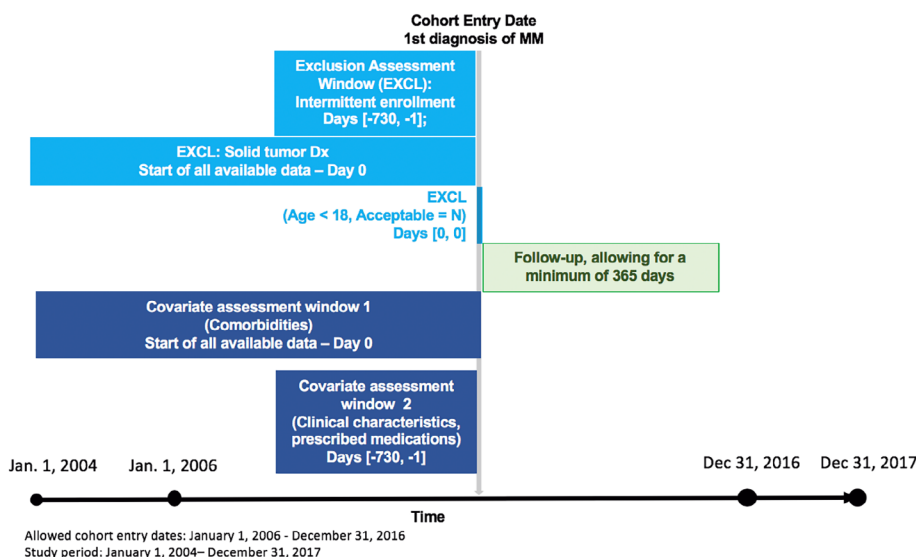


FIGURE 1 Study design [Colour figure can be viewed at wileyonlinelibrary.com]

were small discrepancies in patient distribution across age categories and demographic characteristics (absolute difference range of 0.0% to 1.1%) (Table 1). High agreement was observed for the prevalence of comorbidities (CVD, CKD, gout, hypertension, osteoarthritis, osteopenia, osteoporosis, and RA), with a match rate of 75% and an absolute difference in prevalence of 0.0%–0.3%. Agreement was observed for several clinical characteristics of symptomatic bone pain and SREs (match rate of 50% and 85.7%, respectively). The absolute difference for SREs prevalence ranged from 0.0%–0.2%, with exact agreement for pathological fracture, spinal cord compression, radiation therapy to bone and surgery to bone.

We observed a lower agreement on specific drug exposure (match rate of 28.6% and an absolute difference in prevalence of 0.0%–19.6%). Exact concordance was also observed for laboratory investigations (hypercalcemia, renal impairment, and anemia) and valid test results. The Phase I analysis yielded several discrepancies which were subsequently addressed and resolved in the Phase II analysis by clarifying operational decisions not fully specified in the protocol or SAP.

3.2 | Phase II: Revised analyses (analysis iteration based on Phase I findings)

Following further detailed specification of study choices that led to different interpretations during Phase I, the analysis was repeated and

a 100% match in the number of patients (N = 2646), demographics, bone pain, SREs, drug exposure, laboratory investigations, and valid laboratory results was achieved. A negligible discrepancy remained in the estimated prevalence of CKD (absolute difference 0.1%), as well as osteoarthritis and gout (absolute different 0.0%), driven by the data of an individual patient with multi-morbidity (Figure 2, Table 1).

3.3 | Comparison of two approaches in Phase I and Phase II

A mismatch in aggregated results triggered a cascade of steps to investigate and remediate discrepancies identified during Phase I. As a first step, the two set of independent investigators reviewed the protocol and SAP definitions against implementation and discussed individual interpretations, such as assumptions and algorithms used, followed by a more detailed investigation into time interval specifications, which resolved a majority of the discrepancies. For any remaining discrepancies, the investigators reviewed patient-level data. Amending study documents (e.g., SAP) was an iterative process.

Table 2 displays the sources of the discrepancies in Phase I and the operational decisions utilized in Phase II. The workflow – and transparency during each step – for conducting database studies using routine-collected longitudinal healthcare data begins with the (1) processing of raw data to (2) developing the study design and

TABLE 1 Results from the reproducibility assessment between SAS and AEP in the Phase I (initial analysis) and Phase II (revised analysis)

Broad category	Definition used	Description	Phase 1: Initial analysis		Phase 2: Revised analysis	
			Absolute difference ^a	Match rate ^b AEP versus SAS (%)	Absolute difference ^a	Match rate ^b AEP versus SAS (%)
Observations	READ codes	NDMM patients, n	2	N/A	0	N/A
Demographics	N/A	Age, gender	0.0%–1.1%	50% (1/2)	0.0%	100% (2/2)
Conditions	READ codes	Comorbidities	0.0%–0.3%	75% (6/8) ^c	0.0–0.1%	87.5% (7/8) ^c
	READ codes	Symptoms (bone pain)	0.0%–0.1%	50% (2/4)	0.0%	100% (4/4)
	READ codes	Clinical events (SREs)	0.0%–0.2%	85.7% (6/7)	0.0%	100% (7/7)
Drug exposure	Product codes	Record of specific drug	0.0%–19.6%	28.6% (4/14)	0.0%	100% (14/14)
	Product codes	Absence of drug	10.8%	0% (0/1)	0.0%	100% (1/1)
Procedure	ENTTYPE READ codes	Laboratory investigation and results	0.0%	100% (3/3)	0.0%	100% (3/3)

Abbreviations: AEP, Aetion Evidence Platform; N/A, not applicable; NDMM, newly diagnosed multiple myeloma.

^aAbsolute difference is defined as the range of difference (in forms of either counts or percentages) between the AEP and SAS results within each category.

^bMatch rate, AEP versus SAS, is defined as a percentage calculated as the number of variables with an exact match in the estimated prevalence between the AEP and SAS results, divided by the total number of variables within each category.

^cOne patient had chronic kidney disease, osteoarthritis, and gout. However, since the absolute difference in prevalence for osteoarthritis and gout were 0.0%, those conditions were considered concordant and contributed to the match rate numerator.

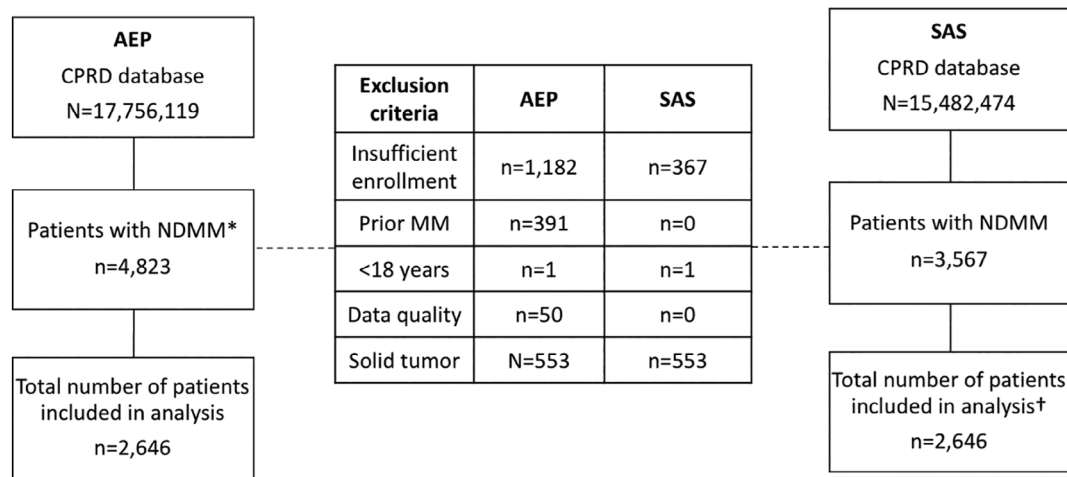


FIGURE 2 Flow diagram of patient inclusion after reconciliation in Phase 2. AEP, Aetion Evidence Platform; CDM, common data model; MM, multiple myeloma; NDMM, newly-diagnosed multiple myeloma; OMOP, Observational Medical Outcomes Partnership; SAS, Statistical Analysis System. Note: Analyses performed using SAS data used CPRD data converted to Observational Medical Outcomes Partnership (OMOP) common data model (CDM); analyses conducted using AEP used the raw data. The AEP analysis applied the CPRD acceptable flag after identifying patients with MM. The former database includes only patients with CPRD data of acceptable quality for research and contains a total of approximately 15 million patients. The raw CPRD data contains approximately 18 million patients, regardless of the data quality. *A total of $n = 391$ patients with NDMM met the cohort entry criteria of first MM diagnosis; however, these $n = 391$ patients were subsequently excluded due to prior MM. †Near-identical number of eligible patients with newly diagnosed MM was achieved in Phase I ($N = 2646$ in AEP vs. $N = 2648$ in SAS). Two patients identified using SAS was due to different assumptions made about which data tables (i.e., clinical, referral, and test tables) to query patients with specific disease diagnoses

operational decisions, and finally (3) deciding on analytical choices. Each step provides opportunities to develop sufficiently clear study documentation (Figure 3).⁴ Lack of specificity in these steps during protocol development and SAP contributed to initial discrepancies.

3.3.1 | Transparency in raw data processing

Our direct replication study highlights the importance of understanding, communicating and documenting data source parameters (e.g., data extraction date, data source range, data cleaning and transformation) in the initial stage of conducting a database study. In the current study, the independent investigators used the same data source (CPRD), the same data cut (June 22, 2018) and the same data range (November 21, 1987 – June 30, 2018); however, different data model conversions and hence different data structures were used. The rapid-cycle analytics tool used raw CPRD data processed with a so-called “adaptive rule system” whereas the SAS analyses were based on CPRD data that had been pre-mapped to the OMOP CDM. There are pros and cons to both data model approaches.²³ The OMOP CDM allows the construct of a standard vocabulary for different medical concepts; however, a preconfigured CDM could potentially result in information loss due to incomplete mapping of construct terms.^{24,25} The remaining negligible discrepancies might have arisen from the small loss of information after applying the CDM.

Despite differences in the timing of data processing to exclude patients with poor quality data from the analysis, both analytic approaches eventually yielded the same number of eligible patients in

the analysis set upon reconciliation of the implementation approach (Figure 2). In the OMOP CDM data, patients and practices that were considered unacceptable for research were omitted during the extraction, transformation, and loading (ETL) process of data mapping and conversion; these analyses were based on acceptable CPRD data for research from the beginning. In contrast, the AEP analyses applied the data quality checks (e.g., “Up To Standard (UTS)” dates and GP practices considered “acceptable for research”) as an exclusion criterion when identifying patients with MM.

3.3.2 | Transparency in study design and operational decisions

In the current study, a common study protocol, implemented by two teams of independent researchers, included an initial study design schema, operational definitions of all variables, and appended code lists and algorithms. A SAP was also co-developed, which included details of the analyses to be undertaken (e.g., handling of missing data). Despite these common study materials capturing the operational and analytical choices, several discrepancies were obtained in Phase I of the study. Subsequent thorough investigations revealed various reasons (Table 2), such as differences in the interpretation of time assessment periods including the eligibility enrollment window.

Eligibility enrollment window

The enrollment window for the inclusion and exclusion criteria is the time window prior to the patient's study entry date, often called the

TABLE 2 Improvement of operational decisions for Phase II following identification of discrepancies observed in Phase I

Variables	Source of discrepancies and match rate <100% in Phase I	Final operational decisions for Phase II
Observations NDMM patients (n)	Realization that patient diagnoses were captured across different data tables (i.e., clinical, referral, and test tables) in CPRD, leading to differing use of data tables by the independent investigators to identify patients with MM	Consensus regarding which data tables to use for identifying patients with MM diagnosis <ul style="list-style-type: none"> • Inclusion of relevant diagnoses from only the Clinical and Referral tables • Non-use of diagnoses in the Test table
Demographics Age Gender ^a	Difference in methods used for handling date data to impute missing age when only year and month of birth are available	Consensus on rules for age calculation: <ul style="list-style-type: none"> • The year of birth (YOB) was defined as YOB + 1800 • If month of birth (MOB) was available, the month of birth was defined as MOB, and the day of birth defined as day 15 of the month • If MOB was not available, then the month and date of birth was defined as June 30th
Conditions Co-morbidities Symptoms (bone pain) Clinical events (SREs)	Lack of specificity in defining the baseline assessment window (e.g., whether or not to include cohort entry date for assessing chronic conditions such as osteoporosis)	Clarification of time assessment windows: <ul style="list-style-type: none"> • Covariate assessment window for comorbidities to include the cohort entry date (Day 0), based on the assumption that the chronic conditions cannot be cured • Covariate assessment window for clinical characteristics (i.e., symptoms, clinical events) to exclude the cohort entry date (Day 0) to assess bone pain and SREs exclusively prior to the MM diagnosis
	Realization that patient diagnoses were captured across different data tables (i.e., clinical, referral, and test tables) in CPRD, leading to differing use of data tables by the independent investigators to identify patients with specific comorbidities	Consensus regarding which data tables to use for identifying patients with specific comorbidities: <ul style="list-style-type: none"> • Inclusion of relevant diagnoses from only the Clinical and Referral tables • Non-use of diagnoses in the Test table
	Differences in data structure of CPRD (i.e., use of transformed data [OMOP CDM structure] in SAS programming), likely contributed to the remaining small discrepancies that were present in comorbidities only	Consensus that the remaining data-driven discrepancies were unresolvable <ul style="list-style-type: none"> • The rapid-cycle analytics tool used raw CPRD and the in-line programming analyses used CPRD data pre-mapped to OMOP CDM
Drug exposure Record of specific drug Absence of drug	Lack of specificity in exposure assessment window (i.e., when is the start date, and whether or not to include cohort entry date in the baseline period)	Agreement on the time assessment window: <ul style="list-style-type: none"> • Covariate assessment window for drug exposures to start from the 730 days prior till 1 day prior to the index date (Day 0) to assess treatment usage exclusively prior to the MM diagnosis
	Lack of a definition for the absence of treatment led to the use of different algorithms by the independent investigators. For instance, patients without any occurrence of analgesic versus patients who did not start any analgesics during the baseline period	Agreement on the definition for absence of treatment: <ul style="list-style-type: none"> • Absence of treatments was measured as not starting any analgesics during the baseline period
Procedures Laboratory investigation	Not applicable; exact concordance in phase I	New decision to facilitate interpretation of the data: <ul style="list-style-type: none"> • For investigation, we included any tests (not requiring valid test only)

^aExact concordance achieved for gender in Phase I.

cohort entry date or index date. Different interpretations of the time assessment periods specified in the protocol and SAP were employed. For example, terms such as “baseline” did not specify, for each covariate, whether the cohort entry date should or should not be included in the assessment period.

3.3.3 | Transparency in analytical choice decisions

Several discrepancies were due to investigator-driven interpretation of variable definitions, algorithms, and the exposure assessment window, which were resolved following alignment between both investi-

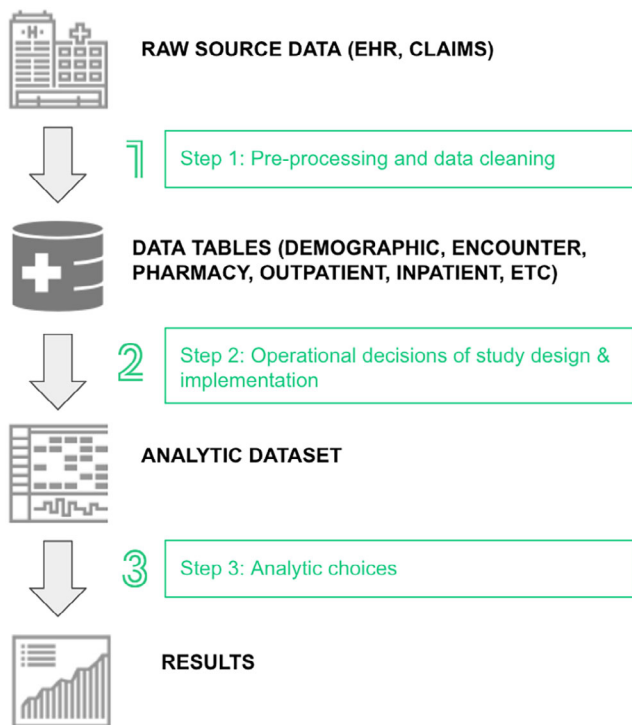


Figure adapted from Wang et al, PDS 2017; 26: 1018-1032.

FIGURE 3 Decisions surrounding the workflow from raw data source to study results [Colour figure can be viewed at wileyonlinelibrary.com]

gators. The remaining discrepancies were data-driven due to differences in data structure and could not be fully addressed in Phase II.

Variable definitions and algorithms

While the variable definitions in the protocol and SAP included a general description of the algorithm and the code list for diagnoses, comorbidities, and clinical conditions, different assumptions were made by the two sets of independent investigators. First, a lack of detail regarding which specific data tables within CPRD to query accounted for the initial discrepancies for all diagnosis events in our study, including MM diagnosis, comorbidities, and baseline clinical conditions. To avoid discrepancies, study documentation should specify both the relevant data fields as well as the data tables to query these fields, depending on the data source used. Second, the inherent logic for certain variable with complicated algorithms was not specific, such as Boolean logic between various components of the algorithms. For example, for events spanning multiple days, such as medication use or hospitalizations, it is important for study documentation to specify whether a patient should be counted as having the medication or event if the event overlaps the time window of interest or only if it begins during the time window of interest.

Exposure assessment window

The exposure assessment window describes the time window for identifying the exposure status. A similar lack of specificity in this time

window accounted for the initial discrepancies between baseline medication use and comorbidities from the two sets of results. While it was specified that baseline medication use would be assessed during the baseline period, one approach included a look back of 2 years and the other of all available data.

4 | DISCUSSION

In this direct replication study characterizing newly-diagnosed MM within a UK-based EHR, after a thorough investigation on the initial discrepancies, we achieved an exact match of the analytic population and nearly 100% match rate using two different analytic approaches, namely a rapid-analytics tool and traditional statistical software. This replication exercise demonstrated that differences in study results are often due to insufficient level of detail in the study protocol or SAP.

4.1 | Standards in transparency and reproducibility

Transparency of the study design and operational decisions shared by the independent investigators were critical to facilitating this direct replication study, highlighting the importance of not only following standard protocol and reporting guidelines (Box 1), but applying greater specificity and transparency to protocol development in the study planning phase. These guidelines on conducting²⁶⁻²⁹ and reporting³⁰⁻³² non-randomized studies have been developed by professional societies, governmental agencies, and pharmacoepidemiological experts. More recently, a ISPOR-ISPE joint task force convened to develop recommendations to ensure RWE can approximate the gold-standard randomized controlled trial and provide “causal conclusions.”³³ The first report from this joint task force provided recommendations for transparency in the “study hygiene” (i.e., planning and procedures) when evaluating treatment and/or comparative effectiveness studies using RWD.³⁴ One recommendation focuses on replication to obtain the same results using the same data and same analytic methods, explaining “Full transparency in design and operational parameters, data sharing, and open access in clinical research will not only increase confidence in the results but will also foster the reuse of clinical data.” The second ISPOR-ISPE joint task force paper⁴ further identifies specific parameters requiring full transparency and explicit reporting of key decisions regarding the data source and study design: data source, overall study design, inclusion/exclusion criteria, exposure definitions, follow-up time, outcome definition, covariate definitions, control sampling, and statistical software. Our findings highlight that second order temporal anchors (such as enrollment window, covariate assessment window, follow-up window, exposure assessment window) should be well specified prior to the analysis being undertaken. In a large-scale reproducibility study of 6 protocols and 31 publications, Wang et al. also identified this lack of clarity on temporal anchors as one of the challenges for replicating source studies.³ Including a standardized visual representation of the study design (such as Figure 1) and key time anchors in the protocol of studies using longitudinal health care databases would preclude such

Box 1 Standard considerations for ensuring transparency and reproducibility of real-world research protocols and findings

- Follow standard protocol^{28,29} and reporting guidelines³⁰⁻³² developed and endorsed by governmental agencies^{26,27} professional societies,^{4,34} and other experts to ensure necessary study parameters are fully described and reported.
- Ensure the protocol and reporting include operational definitions of all variables, including exposures, outcomes, potential confounders; include codes lists and algorithms.^{28,29,31}
- Consider reporting supplemental information that specifies access to the protocol, raw data, or programming codes.³⁰
- Describe statistical methods including how missing data were handled^{28,31} and statistical software program utilized.⁴
- Define critical time periods and temporal anchors in the study design such as study entry date, enrollment windows, washout or lookback periods, exposure assessment windows, and follow-up window.⁴ Standardized design diagrams allow for a quick visual representation of key temporal parameters.^{4,36} Such time periods should be clearly specified and included in study planning documents.
- Understand and fully report pertinent data source parameters such as data extraction date, data source range, data cleaning and transformations when reporting the study methods and findings⁴ (e.g., manuscripts) as well as in study documents (e.g., protocols).

temporal ambiguities.³⁵ While best practices in reporting studies using RWD exist (Box 1), our study highlighted that key parameters regarding data sources and important operational decisions (e.g., the boundaries of the enrollment and time assessment windows) should be explicitly stated during study planning and the protocol development stage.

4.2 | Strengths and limitations

One of the key strengths of our study was the large number of patients newly diagnosed with MM identified in CPRD and the availability of symptoms and specific diagnostic investigations for MM in the CPRD database, allowing us to investigate the reproducibility of these findings. Another strength is the achievement of nearly 100% concordance in the final analysis using the two different analytic approaches. While other studies published in the literature previously conducted reproducibility studies,³ very few studies evaluated the reproducibility among patients newly diagnosed with MM. Ensuring the availability of

a study's data, methods, and documentation, as outlined by Peng et al,⁷ will also facilitate reproducibility in epidemiologic research and improve the confidence of observational study findings. The current study, however, is limited to structured data only and does not cover information captured in free text fields and/or non-automated test results.

5 | CONCLUSION

In conclusion, this direct replication study characterizing patients newly diagnosed with MM within a UK-based EHR database demonstrated that study reproducibility requires maximal transparency at each phase of RWE generation. Following standard protocol and reporting guidelines allowed a rapid-cycle analytics tool to achieve near-exact replication of findings obtained using traditional statistical software. Specificity of the study design and key operational decisions shared by independent teams of investigators were critical to achieve this successful reproducibility.

ETHICAL STATEMENT

The research protocol was reviewed and approved by the Independent Scientific Advisory Committee (ISAC, protocol 18_292).

ACKNOWLEDGMENTS

This project was funded by Amgen. We wish to acknowledge Amanda Patrick for her critical review of the manuscript and Andrew Weckstein for analytical support in this work.

CONFLICT OF INTEREST

The following personal or financial relationships relevant to this manuscript existed during the conduct of the study: JRW, PM, SLR are employees of and hold stock options or equity in Aetion; NPI was an employee of Aetion and held stock options. AS, JM, DN, GK are employees of and hold stock options in Amgen. During the study conduct and reporting, VB and EH were contract workers for Amgen.

REFERENCES

1. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. *Clin Pharmacol Ther.* 2019;106(1):36-39. <https://doi.org/10.1002/cpt.1426>.
2. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence - What is it and what can it tell us? *N Engl J Med.* 2016;375(23):2293-2297.
3. Wang SV, Verpillat P, Rassen JA, Patrick A, Garry EM, Bartels DB. Transparency and reproducibility of observational cohort studies using large healthcare databases. *Clin Pharmacol Ther.* 2016;99(3):325-332. <https://doi.org/10.1002/cpt.329>.
4. Wang SV, Schneeweiss S, Berger ML, et al; joint ISPE-ISPOR Special Task Force on Real World Evidence in Health Care Decision Making. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Pharmacoepidemiol Drug Saf.* 2017;26(9):1018-1032. <https://doi.org/10.1002/pds.4295>.
5. Wang SV, Schneeweiss S. Improving Transparency and Replicability of Healthcare Databases to Increase Credibility of "Real World" Evidence. *BMJ Opinion.* <https://blogs.bmj.com/bmj/2017/12/15/improving-transparency-and-replicability-of-healthcare-databases-to-increase-credibility-of-real-world-evidence/>. December 15, 2017. Accessed November 8, 2019.

6. Franklin JM, Pawar A, Martin D, et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. *Clin Pharmacol Ther*. 2019;107:817-826. <https://doi.org/10.1002/cpt.1633>.
7. Klungel OH, Kurz X, de Groot MC, et al. Multi-centre, multi-database studies with common protocols: lessons learnt from the IMI PROTECT project. *Pharmacoepidemiol Drug Saf*. 2016;25(Suppl 1):156-165. <https://doi.org/10.1002/pds.3968>.
8. Coughlin SS. Reproducing epidemiologic research and ensuring transparency. *Am J Epidemiol*. 2017;186(4):393-394. <https://doi.org/10.1093/aje/kwx065>.
9. Shepherd BE, Blevins Peratikos M, Rebeiro PF, Duda SN, McGowan CC. A pragmatic approach for reproducible research with sensitive data. *Am J Epidemiol*. 2017;186(4):387-392. <https://doi.org/10.1093/aje/kwx066>.
10. Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. *Am J Epidemiol*. 2006;163(9):783-789. <https://doi.org/10.1093/aje/kwj093>.
11. Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol*. 2019;48(6):1740-1740g. <https://doi.org/10.1093/ije/dyz034>. <https://pubmed.ncbi.nlm.nih.gov/30859197/>.
12. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827-836.
13. Berger ML, Curtis MD, Smith G, Harnett J, Abernathy AP. Opportunities and challenges in leveraging electronic health record data in oncology. *Future Oncol*. 2016;12(10):1261-1274.
14. Khozin S, Blumenthal GM, Pazdur R. Real-world data for clinical evidence generation in oncology. *J Natl Cancer Inst*. 2017;109(11):dxj187.
15. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract*. 2010;60(572):e128-e136.
16. Cancer Research UK. Myeloma Statistics. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma#heading-Five>. Accessed October 2019.
17. Rajkumar SV, Dimopoulos MA, Palumbo A, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol*. 2014;15(12):e538-e548.
18. Raluy-Callado M, Lambrelli D, DeCosta L, Gonzalez-McQuire S. Retrospective cohort study using data from the UK Clinical Practice Research Datalink and hospital episode statistics to assess unplanned hospitalisation in patients with multiple myeloma. *Value Health*. 2015;18(7):A672.
19. Shephard EA, Neal RD, Rose P, Walter FM, Litt EJ, Hamilton WT. Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case-control study using electronic records. *Br J Gen Pract*. 2015;65(631):e106-e113. <https://doi.org/10.3399/bjgp15X683545>.
20. Walley T, Mantgani A. The UK general practice research database. *Lancet*. 1997;350(9084):1097-1099.
21. Garcia Rodriguez LA, Perez Gutthann S. Use of the UK general practice research database for pharmacoepidemiology. *Br J Clin Pharmacol*. 1998;45(5):419-425.
22. Fralick M, Kesselheim AS, Avorn J, Schneeweiss S. Use of health care databases to support supplemental indications of approved medications. *JAMA Intern Med*. 2018;178(1):55-63. <https://doi.org/10.1001/jamainternmed.2017.3919>.
23. Schneeweiss S, Brown JS, Bate A, Trifirò G, Bartels DB. Choosing among common data models for real-world data analyses fit for making decisions about the effectiveness of medical products CPT. *Clin Pharm Ther*. 2019;107:827-833. <https://doi.org/10.1002/cpt.1577>.
24. Zhou X, Murugesan S, Bhullar H, et al. An evaluation of the THIN database in the OMOP common data model for active drug safety surveillance. *Drug Saf*. 2013;36:119-134. <https://doi.org/10.1007/s40264-012-0009-3>.
25. Rijnbeek PR. Converting to a common data model: what is lost in translation? Commentary on "Fidelity assessment of a Clinical Practice Research Datalink conversion to the OMOP common data model". *Drug Saf*. 2014;37(11):893-896.
26. U.S. FDA: Guide for Industry and FDA Staff. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. <https://www.fda.gov/files/drugs/published/Best-Practices-for-Conducting-and-Reporting-Pharmacoepidemiologic-Safety-Studies-Using-Electronic-Healthcare-Data-Sets.pdf>. May 2013. Accessed October 22, 2020.
27. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) and European Medicines Agency (EMA). Guide of Methodological Standards in Pharmacoepidemiology (Rev. 2). http://www.encepp.eu/standards_and_guidances/documents/ENCePPGuideofMethStandardsinPE_Rev5.pdf. June 2013. Accessed October 22, 2020.
28. International Society for Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf*. 2016;25:2-10. <https://doi.org/10.1002/pds.3891>.
29. Agency for Healthcare Research and Quality (AHRQ). Publication no. 12(13)-EH099. In: Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, eds. *Developing a Protocol for Observational Comparative Effectiveness Research; A User's GUIDE*. Rockville, MD: Agency for Healthcare Research and Quality; 2013.
30. von Elm E, Altman DG, Egger M, et al; for the STROBE Initiative. STROBE statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147:573-577.
31. Benchimol EI, Smeeth L, Guttmann A, et al; RECORD Working Committee. The REporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLoS Med*. 2015;12:e1001885.
32. Langan SM, Schmidt SAJ, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*. 2018;363:k3532. <https://doi.org/10.1136/bmj.k3532>.
33. Greenfield S. Making real-world evidence more useful for decision making. *Value Health*. 2017;20:1023-1024.
34. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task force on real-world evidence in health care decision making. *Value Health*. 2017;20(8):1003-1008. <https://doi.org/10.1016/j.jval.2017.08.3019>.
35. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical depiction of longitudinal study designs in health care databases. *Ann Intern Med*. 2019;170:398-406. <https://doi.org/10.7326/M18-3079>.
36. Reproducible Evidence: Practices to Enhance and Achieve Transparency (REPEAT). Design Diagram Templates. www.repeatinitiative.org/projects.html. Accessed October 22, 2020.

SUPPORTING INFORMATION

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

How to cite this article: Seesaghur A, Petruski-Ivleva N, Banks V, et al. Real-world reproducibility study characterizing patients newly diagnosed with multiple myeloma using Clinical Practice Research Datalink, a UK-based electronic health records database. *Pharmacoepidemiol Drug Saf*. 2021;30:248-256. <https://doi.org/10.1002/pds.5171>