The BIOMarkers in Atopic Dermatitis and Psoriasis (BIOMAP) glossary: developing a lingua franca to facilitate data harmonization and cross-cohort analyses

DOI: 10.1111/bjd.20587

Dear Editor, The BIOMarkers in Atopic dermatitis and Psoriasis (BIOMAP) is a large European consortium aiming to advance personalized medicine for atopic dermatitis and psoriasis by identifying biomarkers that predict therapeutic response and disease progression. BIOMAP brings together clinicians, researchers, patient organizations and pharmaceutical industry partners, and encompasses data from over 60 individual studies, including randomized clinical trials, population-based cohorts and deeply phenotyped disease registries. The curation and harmonization of data and biosamples from these established studies will facilitate cross-cohort clinical and molecular analyses, increasing the potential to identify small-effect estimates and to better stratify disease subtypes. This research letter serves to disseminate BIOMAP's pathway to data harmonization and will inform future collaborative research endeavours.

Pooling data from diverse studies presents inherent challenges. Each study has different methodologies, research objectives and outcomes. Data harmonization improves the comparability of existing studies by converting similar variables to a common format and creating 'harmonized datasets', which can be used for cross-cohort analyses. Figure 1 outlines how BIOMAP follows existing data harmonization guidelines,[1] ensuring that clinically appropriate and meaningful conclusions can be drawn.

BIOMAP's objectives were outlined in the project proposal (step 0). During protocol development, a list of variables pertinent to BIOMAP's key research questions was devised. These predefined 'BIOMAP categories' included clinical phenotypes, disease associations, environmental/lifestyle factors, treatments and outcome measures. Next, a detailed mapping exercise was performed to explore what data were available in a subset of the studies underpinning BIOMAP. This involved the custodians of individual study datasets assigning a BIOMAP category to each variable in their study's data dictionary. Annotated data dictionaries were assimilated

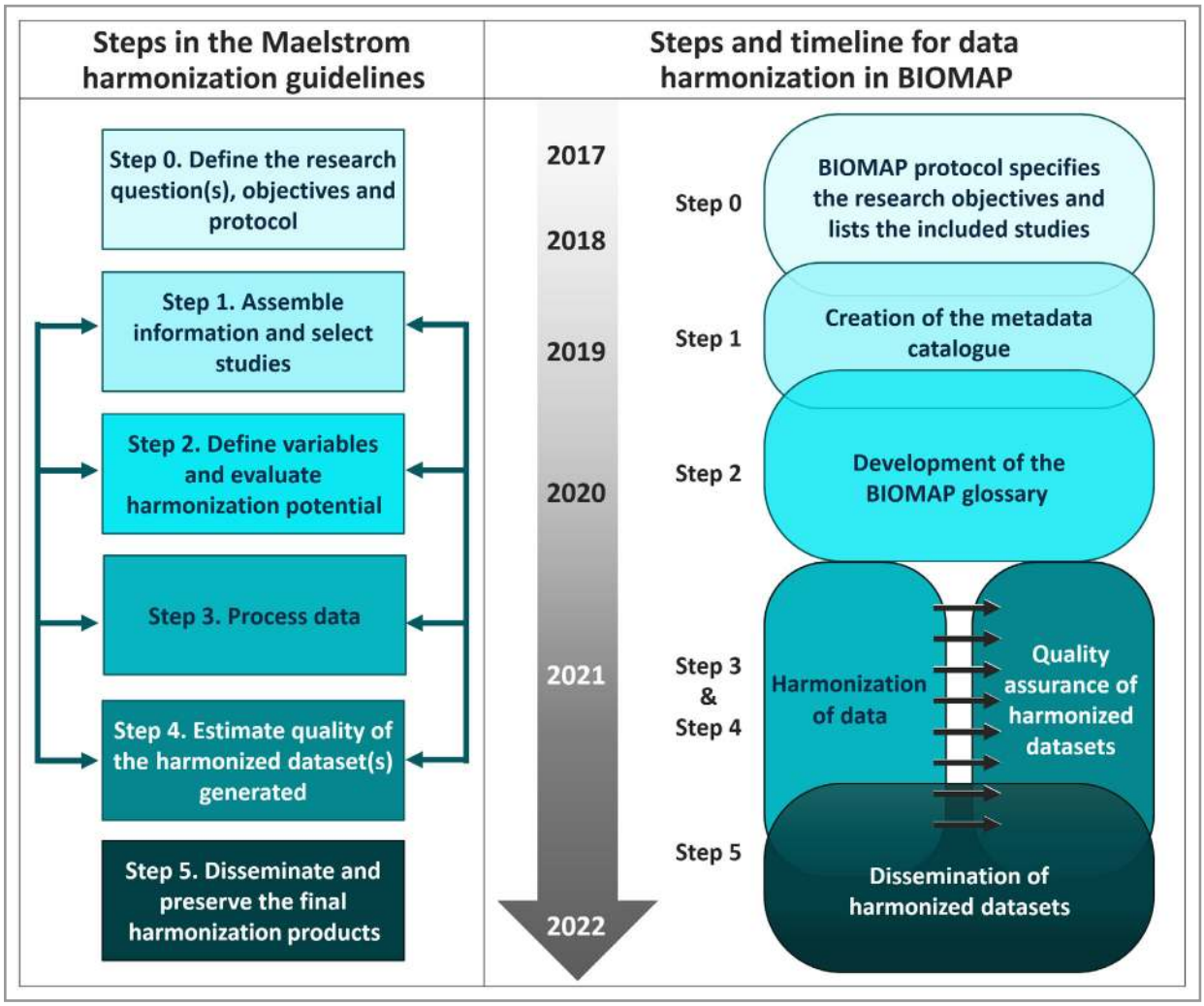

Figure 1 The pathway to data harmonization of BIOMarkers in Atopic dermatitis and Psoriasis (BIOMAP) studies. (Left) Proposed steps for retrospective data harmonization (adapted from the Maelstrom guidelines).[1] (Right) Implementation of these steps for data harmonization in BIOMAP. Overlapping boxes represent steps running concurrently. Following finalization of the BIOMAP glossary (step 2), harmonization of individual study datasets started in a pragmatic and prioritized manner, based on the availability of data and proposed cross-cohort analyses. Quality assurance (step 4) is integrated with step 3 in our harmonization pipeline, expediting the availability of harmonized datasets for cross-cohort analyses.

into a clinical 'metadata catalogue' indexed according to the BIOMAP categories, generating a high-level overview of the clinical variables recorded in this sample of BIOMAP studies (step 1). The metadata catalogue identified similarities and discrepancies between studies, and formed the foundation of the BIOMAP glossary.

The BIOMAP glossary defines a list of core variables, using harmonized terminology and data format (step 2), and will be used to create harmonized datasets. The Glossary Development Team comprised clinical, bioinformatics, biostatistics and laboratory expertise, and discussed the potential contents of the glossary (11 members, representing five BIOMAP organizations). Discussions were informed by the metadata catalogue, literature reviews and existing harmonization initiatives, including the TREatment of ATopic eczema (TREAT) Registry Taskforce,[2] Harmonising Outcome Measures for Eczema[3] and the International Psoriasis Council.[4]

A BIOMAP webinar introduced data harmonization to the wider BIOMAP consortium, illustrating the fundamental role the glossary would play in downstream BIOMAP analyses. Following the webinar, glossary stakeholders were identified ($n = 67$, including work-package leaders, dataset custodians, clinicians and analysts from 28 BIOMAP organizations).

A draft glossary was circulated to the glossary stakeholders who refined and approved the finalized glossary through a series of three interactive Zoom meetings. Following group discussion, any amendments to the proposed glossary were approved or rejected through anonymous polling, using in-built Zoom functionality (30 polls). The outcome of voting was accepted

with a simple majority (median agreement 100%; range 57–100) and the BIOMAP glossary version 1.0 was finalized.

Primary datasets are being transformed to conform to the content and structure of the BIOMAP glossary, creating harmonized datasets (step 3). Iterative discussions between each dataset custodian and the harmonization bioinformaticians culminate with a dataset-specific mapping document specifying how individual variables will be transformed to the glossary-defined dataset, thus ensuring accurately harmonized data (step 4). Harmonized datasets are made available on a secure, centralized and access-controlled data platform (step 5). Harmonized clinical datasets complement a carefully curated bioresource of archived and newly obtained biospecimens, which will be used for multiomic profiling of skin and blood.

The structure of the BIOMAP glossary was inspired by the internationally recognized Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).[5] The OMOP CDM adopts existing vocabularies, such as SNOMED Clinical Terms,[6] and was developed to implement standardized analytical approaches on large observational datasets. During glossary development, deviations from the OMOP CDM were made where existing variables were not represented in the OMOP-defined terminology or where dermatological research required additional granularity (e.g. detailed information regarding phototherapy). The OMOP CDM tabular structure was adjusted to match BIOMAP analysts' requirements. Full compatibility with the OMOP CDM is a priority for further development of the glossary.

The publicly available BIOMAP glossary may benefit investigators beyond the BIOMAP consortium who could prospectively align future studies with the glossary's clinical variables, thus facilitating comparative analyses.[7] Published dermatological research using OMOP approaches is currently limited.[8] Cooperation between BIOMAP and OMOP, leading to the incorporation of BIOMAP customizations into the OMOP CDM is an appealing prospect. Collaboration could further enhance the potential for dermatological research using large observational datasets.

Acknowledgments: We would like to thank the following people who formed the BIOMAP Glossary Stakeholder Group and contributed to the development of the BIOMAP glossary: D. Aagaard-Hansen, A. Aguayo Orozco, I.A. Barbosa, K. Bønnelykke, N. Boraczynski, S.J. Brown, A. Budu-Aggrey, P. Bryce, S. Maria Christensen, J. Di Domizio, K. Dormanns, S. Eyerich, A. Farag, A. Federico, N. Forgó, V. Fortino, S. Gerdes, M. Gilliet, M. Gossel, D. Greco, M. Griffiths, B. Hinrichs, M. Hjelmsoe, N. Højsgaard Jørgensen, B. Homey, Y. Kelkar, E. Kieras, S. Langan, T. Litman, W. Rush Koopmann, D. McKean, L. Moebus, L. Moitinho e Silva, A. Napieralski, P. Oláh, K. Page, K. Polychronopoulos, M. Reiger, M. Rissetto, H. Sinkko, K. Skak, T. Skoog, M. Soverini, J. Stokholm, C. Traidl-Hoffmann, T. Traks, E. van den Bogaard and K. Vierlinger [full names and affiliations are available in Appendix S1 (see Supporting Information)].

C. Broderick,[1] N. Christian,[2] C. Apfelbacher,[3] A.L. Bosma,[4] N. Dand,[5] S. Ghosh,[6] N. Hangel,[7] M. Hübenthal,[8] M.A. Middelkamp-Hup,[4] J.L. Min,[9,10] A.H. Musters,[4] L. Paternoster,[9,10] E. Rodríguez,[8] V. Satagopam,[6] P. Scordis,[11] P.I. Spuls,[4] S. Szymczak,[12] S. Weidinger,[8] C.H. Smith,[1] and C. Flohr[1] and on behalf of the BIOMAP consortium

[1]*St John's Institute of Dermatology, Guy's and St Thomas' NHS Foundation Trust and King's College London, London, UK;* [2]*ITTM S.A., Information Technology for Translational Medicine, Esch-sur-Alzette, Luxembourg;* [3]*Institute of Social Medicine and Health Systems Research, Otto von Guericke University Magdeburg, Magdeburg, Germany;* [4]*Department of Dermatology, Amsterdam Public Health, Infection and Immunity, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands;* [5]*Department of Medical and Molecular Genetics, School of Basic and Medical Biosciences, Faculty of Life Sciences and Medicine, King's College London;* [6]*Bioinformatics Core, Luxembourg Centre For Systems Biomedicine (LCSB), University of Luxembourg;* [7]*Institute of History and Ethics in Medicine, Technical University of Munich;* [8]*Department of Dermatology and Allergy, University Hospital Schleswig-Holstein;* [9]*MRC Integrative Epidemiology Unit, University of Bristol;* [10]*Population Health Sciences, Bristol Medical School, University of Bristol;* [11]*UCB Pharma (UCB Celltech Ltd); and* [12]*Institute of Medical Biometry and Statistics, University of Lübeck, University Hospital Schleswig-Holstein Campus Lübeck, Lübeck,*

Correspondence: Carsten Flohr.
Email: carsten.flohr@kcl.ac.uk

A full list of author affiliations is provided in Appendix S2 (see Supporting Information).

References

1 Fortier I, Raina P, Van den Heuvel ER *et al.* Maelstrom Research guidelines for rigorous retrospective data harmonization. *Int J Epidemiol* 2016; **46**:103–5.
2 TREAT Registry Taskforce. Available at: https://treat-registry-taskforce.org/ (last accessed 14 June 2021).
3 Harmonising Outcome Measures for Eczema. Available at: http://www.homeforeczema.org (last accessed 14 June 2021).
4 International Psoriasis Council. Available at: https://www.psoriasiscouncil.org/ (last accessed 14 June 2021).
5 Hripcsak G, Duke JD, Shah NH *et al.* Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015; **216**:574–8.
6 SNOMED International. Available at: http://www.snomed.org/ (last accessed 14 June 2021).
7 The BIOMAP Consortium. BIOMAP glossary. Available at: https://zenodo.org/record/4746584#.YPanEy1Q1MM (last accessed 14 June 2021).
8 Sivesind TE, Runion T, Branda M *et al.* Dermatologic Research Potential of the Observational Health Data Sciences and Informatics (OHDSI) Network. *Dermatology* 2021. https://doi.org/10.1159/00514536.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Acknowledgements.

Appendix S2. Full list of author affiliations.

Appendix S3. Author conflicts of interest.

Funding sources: BIOMAP has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 821511. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. L.P. and J.L.M. work in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol, which is supported by the Medical Research Council (MC_UU_00011/1 and MC_UU_00011/5) and the University of Bristol.

Conflicts of interest: Appendix S3 (see Supporting Information).