

Predicting the global distribution of *Burkholderia pseudomallei* and burden of melioidosis

Direk Limmathurotsakul^{1,2,3}, Nick Golding¹, David AB Dance^{4,5}, Jane Messina^{1,6}, David M Pigott¹, Catherine Moyes¹, Dionne B Rolim⁷, Eric Bertherat⁸, Nicholas PJ Day^{2,5}, Sharon J Peacock^{2,9}, Simon I Hay^{1,10}

¹ Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

² Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, 10400, Thailand.

³ Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok, 10400, Thailand.

⁴ Lao Oxford-Mahosot Hospital-Wellcome Trust Research Unit, Vientiane, Lao PDR.

⁵ Centre for Tropical Medicine and Global Health, University of Oxford, Old Road Campus, Oxford, OX3 7FZ, United Kingdom.

⁶ Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom

⁷ Universidade de Fortaleza, Fortaleza, 60811-905, Brazil.

⁸ Department of Pandemic and Epidemic Diseases, World Health Organization, Geneva, 27, Switzerland.

⁹ Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 0QQ, United Kingdom.

¹⁰ Institute of Health Metrics and Evaluation, University of Washington, Seattle, USA

The database for the manuscript is publicly available online as a comma-delimited file for ease of use and the ability to import into a variety of software programs. Each of the 22,338 rows represents a single occurrence record of either human melioidosis case, animal melioidosis case or presence of *B. pseudomallei* in the environment. The fields contained in the database are as follows:

1. OCCURRENCE_ID: Unique identifier for each occurrence in the database after temporal and locational standardisation
2. OCCURRENCE_TYPE: Whether the record represents human melioidosis case, animal melioidosis case or presence of *B. pseudomallei* in the environment
3. OCCURRENCE_EXPORT: Whether the record represents the export event
4. SOURCE: The details of the source
5. YEAR: The year of occurrence
6. REGION: The region where the occurrence lies – values are East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, South Asia and Sub-Saharan Africa.
7. COUNTRY: The name of the country where the occurrence lies
8. LOCATION: The name of the location where the occurrence lies
9. LAT: The latitudinal coordinate of the point or polygon centroid (WGS 1984 Datum)
10. LONG: The longitudinal coordinate of the point or polygon centroid (WGS 1984 Datum)