

Harnessing Individual Participant Data from Clinical Trials to Predict Relapse in Visceral Leishmaniasis



James Patrick Wilson
Green Templeton College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Clinical Medicine

Hilary 2026

Abstract

Abstract goes here.

Acknowledgements

Personal

Acknowledgements go here. Supervisors, colleagues, friends, family. Also need to acknowledge publications including any shared authorship.

Institutional

Can include financial support here.

Contents

List of Abbreviations	vii
1 Introduction	1
2 Background	2
2.1 Introduction	2
2.2 Visceral Leishmaniasis	3
2.2.1 Epidemiology	3
2.2.2 Pathophysiology	10
2.2.3 Clinical Features	11
2.2.4 Diagnosis	12
2.2.5 Treatment	13
2.2.6 Elimination Efforts	14
2.3 Relapse	16
2.3.1 Burden of Relapse	16
2.3.2 Relapse vs. Reinfection	17
2.3.3 Relapse Determinants	17
2.4 Summary	27
3 Systematic Review of Prognostic Models	29
3.1 Introduction	30
3.2 Methodology	32
3.3 Results	35
3.4 Discussion	44
3.4.1 Prognostic Model Assessment	44
3.4.2 Predictors of Mortality	47
3.4.3 Implications for Future Research	48
3.4.4 Limitations	49
3.4.5 Conclusion	50

4 Methodology	51
4.1 Introduction	51
4.2 Data Harmonisation	53
4.2.1 Data Acquisition	54
4.2.2 Data Curation	54
4.2.3 Population at Risk	56
4.2.4 Outcome	59
4.3 Model Development	60
4.3.1 Sample Size	60
4.3.2 Candidate Predictors	62
4.3.3 Descriptive Analysis	66
4.3.4 Missing Data	67
4.3.5 Model Specification	70
4.3.6 Variable Selection	71
4.3.7 Model Performance	72
4.4 Internal Validation	74
4.5 Summary	75
5 Results: Indian Subcontinent Models	76
5.1 Descriptive Analysis	76
5.1.1 Study Characteristics	76
5.1.2 Patient Characteristics	84
5.1.3 Univariable Associations	90
5.1.4 Missing Data	92
5.2 Model Results	93
5.2.1 Model Specification and Coefficient Estimates	93
5.2.2 Model Performance and Internal Validation	97
5.3 Summary	99
6 Results: East Africa Models	102
6.1 Descriptive Analysis	102
6.1.1 Study Characteristics	102
6.1.2 Patient Characteristics	110
6.1.3 Univariable Associations	116
6.1.4 Missing Data	118
6.2 Model Results	119
6.2.1 Model Specification and Coefficient Estimates	119
6.2.2 Model Performance and Internal Validation	122
6.3 Summary	123

7 Discussion	126
7.1 Introduction	126
7.2 Interpretation of Predictors and Causal Plausibility	127
7.2.1 Indian Subcontinent	127
7.2.2 East Africa	127
7.3 Implementation Considerations	127
7.3.1 Indian subcontinent	127
7.3.2 East Africa	127
7.4 Strengths	127
7.5 Limitations	127
7.6 Future Research	128
7.7 Conclusion	128

Appendices

A Appendix — Background	130
A.1 Background Literature Search	130
B Appendix — Model methodology	132
C Appendix — Results: Indian Subcontinent Model	133
D Appendix — East Africa model results	158
References	182

List of Abbreviations

ABD	Amphotericin B deoxycholate
ABLE	Amphotericin B lipid emulsion
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AUC	Area under the receiver operating characteristic curve
BPKIHS	B.P. Koirala Institute of Health Sciences
BMGF	Bill & Melinda Gates Foundation
BMI	Body mass index
CD	Cluster of differentiation
CDISC	Clinical Data Interchange Standards Consortium
CI	Confidence interval
CL	Cutaneous leishmaniasis
CRP	C-reactive protein
Crt	Creatinine
DAT	Direct agglutination test
DND<i>i</i>	Drugs for Neglected Diseases initiative
EC	European Commission
EoT	End of treatment
ESR	Erythrocyte sedimentation rate
GDPR	General Data Protection Regulation
HAART	Highly active antiretroviral therapy
Hb	Haemoglobin
HR	Hazard ratio
HIV	Human immunodeficiency virus
ICT	Immunochemical test

IDDO	Infectious Diseases Data Observatory
IFNγ	Interferon gamma
IM	Intramuscular
IPD	Individual participant data
IQR	Interquartile range
IRS	Indoor residual spraying
ISC	Indian subcontinent
ITN	Insecticide treated nets
IV	Intravenous
KAEP	Kala-azar Elimination Programme
KAMRC	Kala-azar Medical Research Centre
kDNA	Kinetoplast deoxyribonucleic acid
LAMB	Liposomal amphotericin B (AmBisome [®] ; Gilead Sciences)
LEAP	Leishmaniasis East Africa Platform
LST	Leishmanin skin test
MA	Meglumine antimoniate
MCL	Mucocutaneous leishmaniasis
MF	Miltefosine
MPS	Mononuclear phagocyte system
MSF	Médecins Sans Frontières
NCVBDC	National Centre for Vector Borne Diseases Control
NTD	Neglected tropical disease
PCR	Polymerase chain reaction
PD1	Programmed cell death protein-1
PICOTS	Population, Index Model, Comparator Model, Outcome, Timing, Setting
PK-PD	Pharmacokinetic-pharmacodynamic
Plt	Platelets
PM	Paromomycin
qPCR	Quantitative polymerase chain reaction
RBC	Red blood cell

RDT	Rapid diagnostic test
RMRIMS	Rajendra Memorial Research Institute of Medical Sciences
SDA	Single dose liposomal amphotericin B
SDTM	Study Data Tabulation Model
SSG	Sodium stibogluconate
TDR	UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases
TGFβ	Transforming growth factor beta
UN	United Nations
UNDP	United Nations Development Programme
UNICEF	United Nations International Children's Emergency Fund
VL	Visceral leishmaniasis
WBC	White blood cells
WHO	World Health Organization

1

Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus lacinia odio vitae vestibulum vestibulum. Cras venenatis euismod malesuada. Nullam ac erat ante. Sed vel urna at dui iaculis gravida. Praesent euismod, nisl vel tincidunt luctus, nunc urna aliquet nunc, quis aliquam nisl nunc eu lectus. Suspendisse potenti. In hac habitasse platea dictumst.

2

Background

Contents

2.1	Introduction	2
2.2	Visceral Leishmaniasis	3
2.2.1	Epidemiology	3
2.2.2	Pathophysiology	10
2.2.3	Clinical Features	11
2.2.4	Diagnosis	12
2.2.5	Treatment	13
2.2.6	Elimination Efforts	14
2.3	Relapse	16
2.3.1	Burden of Relapse	16
2.3.2	Relapse vs. Reinfection	17
2.3.3	Relapse Determinants	17
2.4	Summary	27

2.1 Introduction

This chapter provides important context for the development and validation of prognostic models predicting VL relapse in the Indian subcontinent (ISC) and East Africa. It begins with an overview of VL, outlining its epidemiology, pathophysiology, clinical features, and management. Emphasis is placed on the epidemiology of VL and the ongoing WHO-supported elimination programmes in the ISC and East Africa, defining the public health landscape in which a VL prognostic model would be implemented.

The section on relapse is presented as a narrative synthesis informed by a systematic review of the literature (full methods in Appendix A). Using pre-specified inclusion criteria, the review identifies and evaluates studies describing the burden, mechanisms, and determinants of relapse. The discussion highlights how relapse constitutes a high-risk infection reservoir that, together with PKDL and VL/HIV co-infection, poses a major threat to elimination efforts — a threat that could be mitigated through early identification of patients at high risk of relapse, enabling timely diagnosis and treatment.

2.2 Visceral Leishmaniasis

The leishmaniases are a diverse group of neglected tropical diseases (NTDs) caused by protozoan parasites of the *Leishmania* genus and transmitted between susceptible mammalian hosts via the bite of infected female sandflies[1]. At least 20 *Leishmania* species infect humans, causing four principal disease forms: cutaneous leishmaniasis (CL), mucocutaneous leishmaniasis (MCL), visceral leishmaniasis (VL), and post kala–azar dermal leishmaniasis (PKDL) (see Box 2.1)[1–4]. Disease form is largely determined by parasite species and strain, resulting in a spectrum of clinical presentations ranging in severity from the relatively common and usually self-healing skin lesions seen in CL, to the often disfiguring mucosal destruction of MCL, and life-threatening systemic illness of VL. PKDL is a disseminated dermal eruption that may occur following VL recovery, and while patients are systemically well, they have been shown to be infective to sandflies[3].

VL represents the most severe manifestation of leishmaniasis, accounting for the vast majority of its morbidity and mortality. The disease overwhelmingly affects impoverished rural populations with poor access to healthcare. VL is also recognised as an opportunistic infection in patients living with human immunodeficiency virus (HIV), in whom co-infection leads to particularly high rates of treatment failure and death.

2.2.1 Epidemiology

This section provides an overview of the current global distribution, transmission dynamics, and risk factors for developing visceral leishmaniasis, highlighting major regional differences in disease burden and risk.

VL is endemic¹ in at least 80 countries across tropical, semi-tropical and temperate climates. The disease is caused by two closely related *Leishmania* species whose

¹Defined by the WHO as the occurrence of at least one autochthonous case with demonstrated local transmission within a country[5].

Box 2.1: Principal disease forms of Leishmaniasis[2, 4]

Cutaneous leishmaniasis (CL) Results in lesions on exposed skin that can lead to ulceration and life-long scarring. Often self-healing within a year, but can manifest atypical and disseminated forms, especially in immunocompromised patients. Up to 1 million new cases per year with most cases occurring in the Americas, Mediterranean basin, Middle East and Central Asia.

Mucocutaneous leishmaniasis (MCL) Rare complication of CL seen especially in the Americas with most cases reported in Bolivia, Perú, and Brazil. Results in destructive ulceration of the oral and nasal mucosa. Highly stigmatising and challenging to treat.

Visceral leishmaniasis (VL) Also known as kala-azar, the most severe form of leishmaniasis caused by *L. donovani* in the Old World and *L. infantum* in the New World. With an estimated 50,000–90,000 cases/year, VL presents with progressive weight loss, splenomegaly and fever.

Post kala-azar dermal leishmaniasis (PKDL) Benign macular and/or papular rash often including the face, arms, and trunk. Affects 5–20% of patients months to years after successful initial treatment for VL in the Indian subcontinent and East Africa. Rarely seen with *L. infantum* infections. Often self-limiting, although known to be infective to sandflies and therefore acts as a disease reservoir.

distribution defines the four principal global regions of high endemicity (see Figure 2.1): *L. donovani*, responsible for anthroponotic transmission (human→sandfly→human) in the ISC and East Africa, and *L. infantum*², responsible for zoonotic transmission (mammal→sandfly→human) in the Americas (limited to Central and South America) and the Mediterranean basin, extending into the Middle East and Central Asia[5, 6]. Together, these two species comprise the *L. donovani* complex.

²Previously referred to as *L. chagasi* in the Americas.

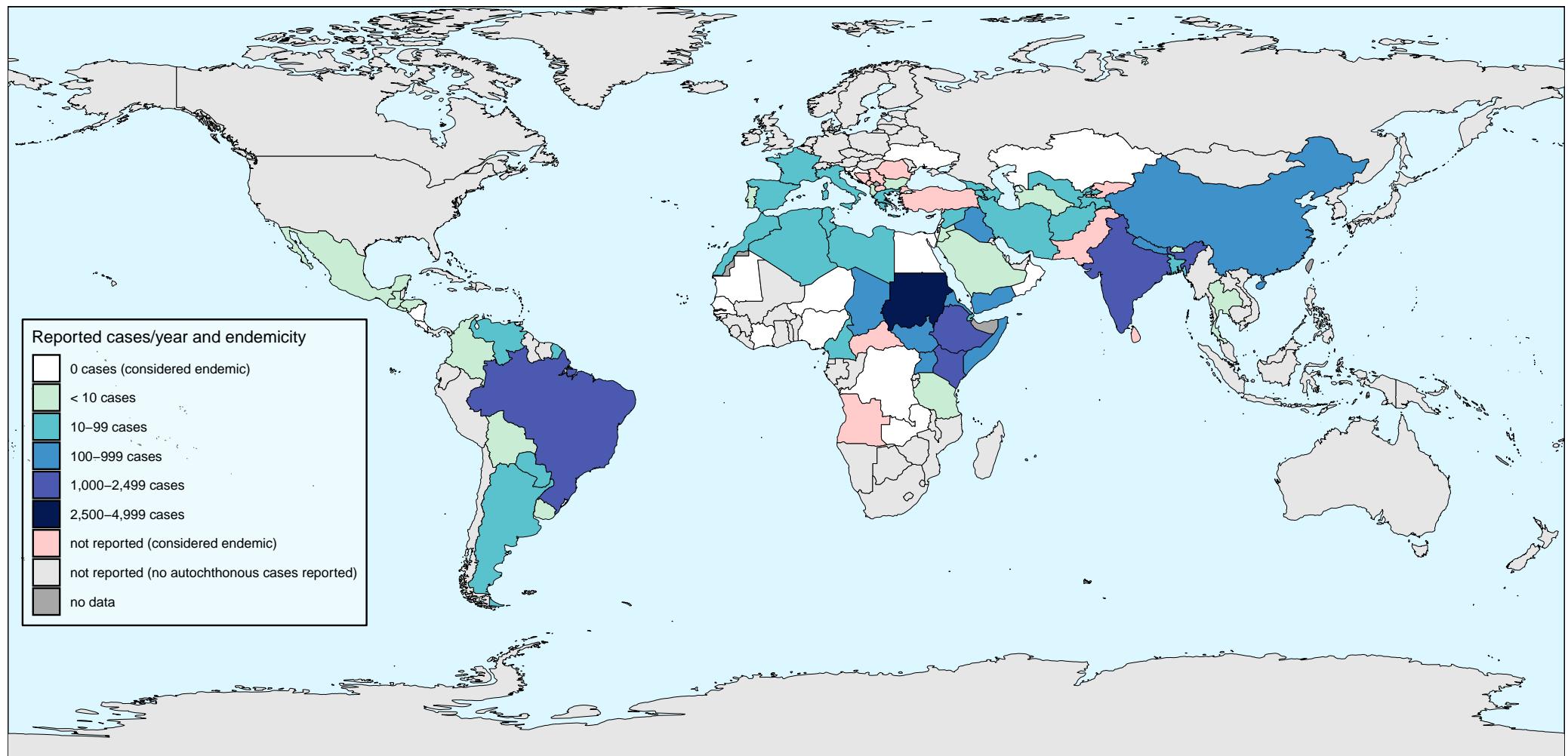


Figure 2.1: Geographical distribution of VL reported per country. Incidence is derived by averaging the reported cases/year over the last 6 reporting years (2019–2024). ‘Considered endemic’: at least one autochthonous case has been reported, with or without the entire cycle of transmission being demonstrated. Data source: World Health Organization Global Health Observatory, accessed 18 January 2026[6].

Disease Burden

Estimating the global burden of VL is problematic. True case numbers are obscured by significant underreporting due to limited access to healthcare, inadequate diagnostic facilities, misdiagnosis, and poor surveillance systems in many endemic countries[7–10]. In 2012, Alvar et al published the results of a WHO-led update to the global incidence of leishmaniasis using country-level reporting from the mid-late 2000s[11]. Underreporting rates were estimated through consultation with country representatives and disease experts. The global incidence was estimated at 200,000–400,000 cases/year, with approximately 80% of the burden originating from the ISC and 15% from East Africa. Compared to official reporting over the same period, this reflected a global underreporting rate of 3.5–7-fold.

Since the publication of Alvar et al, the number of cases has undisputedly fallen, driven largely by reductions in the ISC following the launch of the Kala-Azar Elimination Programme (KAEP) in 2005. Reported incidence dropped from > 50,000 cases/year prior to 2012, to approximately 22,500 cases/year in 2017, and remaining between 11,000 and 13,000 cases/year from 2020–2024 (inclusive)[6] (high incidence country breakdowns presented in Figure 2.2). Notably, this downward trend has persisted despite improvements in surveillance and reporting systems in many endemic countries[5]. Reflecting these changes, the WHO revised its estimated annual incidence in 2017 to 50,000–90,000 new cases/year³[4, 12].

Based on the most recent reporting data from 2024[6], the five countries with the highest case numbers are now, in descending order, Sudan, Kenya, Ethiopia, Brazil, and South Sudan, collectively comprising 78.1% of the global total. In stark contrast to the situation 20 years ago — when the majority of the global burden originated from the ISC — these countries now contribute only 5.5% of all reported cases: India with 464 cases (3.6%), Nepal with 216 cases (1.7%), and Bangladesh with just 23 cases (0.2%).

Mortality

The disease is widely described as ‘fatal without treatment’[2, 3]. Supporting this statement are the high mortality figures recorded during conflict-related epidemics in East Africa over the last 40 years, and prior to effective therapy, 19th century accounts of outbreaks devastating communities across the Gangetic plains[13–15]. Despite

³According to the online WHO Leishmaniasis Fact Sheet: <https://www.who.int/news-room/fact-sheets/detail/leishmaniasis>. Alternate WHO online content estimates 30,000 new cases/year since 2020: <https://www.who.int/health-topics/leishmaniasis> (online material accessed 18 January 2026).

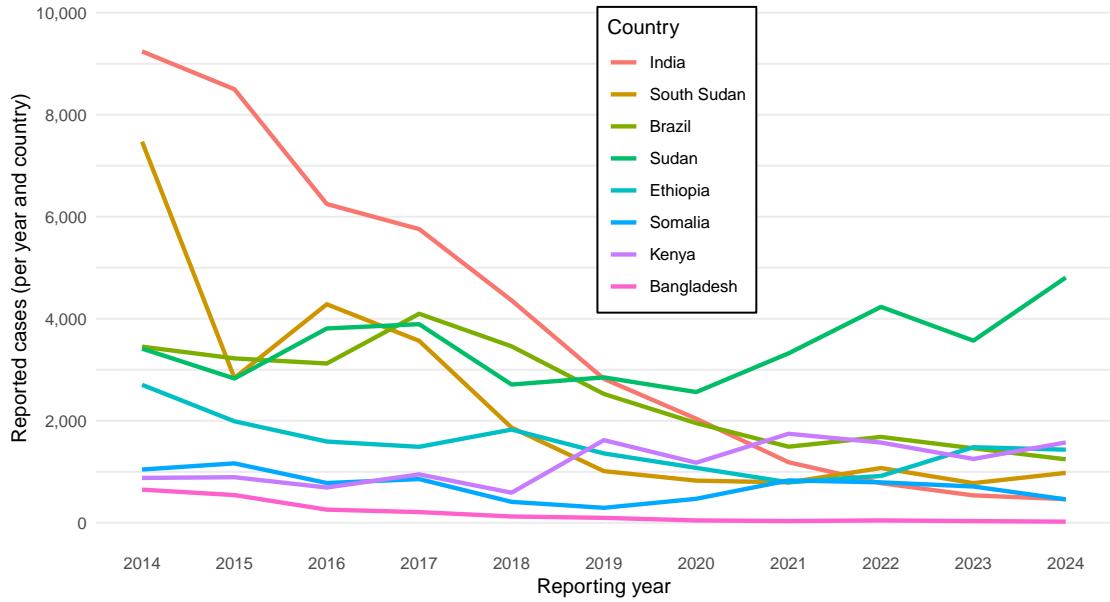


Figure 2.2: Temporal trends of visceral leishmaniasis incidence by country. Comparison between the top 8 countries with the highest average case numbers reported between 2014 and 2024. Data sourced from the World Health Organization Global Health Observatory (accessed 18 January 2026)[6].

this, subclinical disease forms have been reported with spontaneous resolution, although their extent and epidemiological significance remains elusive[16, 17].

With treatment, ~5–15% of cases result in death, although accurate estimates are challenged by a lack of reporting. Where deaths are reported, they frequently only reflect hospital deaths and omit those where a definite diagnosis was missed. In the 2000s, it was often cited that VL was responsible for up to 50,000 deaths/year — the second highest mortality from a parasitic disease after malaria[7] — although more recent estimates tentatively place the figure at a more modest 4,627 deaths/year with a wide uncertainty range of 1,853–8,725 deaths/year[18].

Vector

Measuring 2–4 mm and covered in dense hairs, phlebotomine sandflies (Diptera: Psychodidae) appear distinctly fuzzy under magnification. Females from an estimated 31 species across two genera are known to transmit the parasite between human hosts: *Lutzomyia* in the New World and *Phlebotomine* in the Old World[19]. Sandflies occupy a wide range of ecological niches, found on every continent except Antarctica. Biting occurs from dusk to dawn, with females requiring a blood meal for larval development. During the day they are found in cool and sheltered locations, such as in cracks and crevices in walls as seen with *Ph. argentipes*, responsible for

transmission in the ISC. In East Africa, three sandfly vectors have been implicated in *L. donovani* transmission, defining two distinct and non-overlapping ecological settings: (i) the *Acacia-Balanites* and black cotton soil savannah regions in northern focus, incorporating Sudan, northern South Sudan, and northern Ethiopia, where *Ph. orientalis* thrives, and (ii) the savannah and forested areas in the southern focus, incorporating southern Ethiopia, Kenya, and Uganda, where *Ph. martini* and *Ph. celiae* are seen in association with *Macrotermes* termite mounds.

In addition to sandflies, needle sharing among people who inject drugs was considered an important route of transmission in the southern Mediterranean region during the 1990s and 2000s, particularly among people living with HIV [20]. Exceptionally, transmission can also result from blood transfusion, organ transplantation, laboratory accidents[2], mother-to-child transmission[21], and possibly even sexual contact[22, 23].

Reservoirs

Similar to the majority of *Leishmania* spp. causing CL and MCL, *L. infantum* demonstrates zoonotic transmission, with domestic dogs being the main reservoir host in both the Americas and the Old World. This being said, an ever-increasing list of wild and domestic animals are known to harbour the parasite, including cats, foxes, horses, rodents, bats and opossums, although their relevance to human infection is unclear[24, 25]. An outbreak near Madrid (2009–2012) was attributed to hares[26].

In contrast to *L. infantum*, and importantly from an elimination perspective, *L. donovani* transmission in the ISC and East Africa is predominantly anthroponotic. Xenodiagnosis studies confirm that patients with active VL and those with PKDL are competent human reservoirs, in contrast to individuals with asymptomatic infection[27, 28]. Although *L. donovani* infections have been detected in several animal species in both regions—including cattle, dogs, and rats—their relevance as sources of human transmission remains unproven[29, 30].

Risk Factors

From population prevalence studies we know that only a minority of people with detectable parasites develop symptoms[3, 31]. Risk factors for acquiring an initial asymptomatic infection and subsequent progression to symptomatic disease reflect a tangled ecology of determinants linking host factors (sandfly exposure, immunity, genetics) and parasite factors (strain, virulence, inoculum). A common theme, woven into many of these determinants, is poverty.

In both the ISC and East Africa the median age of infection during stable transmission is similar at 15–20 years. More men than women are infected and develop disease, likely reflecting their increased occupational exposure to sandflies (for example, cattle herding and other outdoor activities)[5].

In the ISC⁴, VL endemicity is centred on the fertile and low-lying alluvial plains of the Ganges river, where high humidity, heavy monsoon rains, and abundant vegetation provide ideal conditions for sustained transmission between sandflies and humans[32]. Significant clustering of cases is seen across the rural farming communities of Bihar, Jharkhand, Uttar Pradesh, and West Bengal States in northeastern India, central and western Bangladesh, and the southeastern plains of Nepal neighbouring India. In a systematic review by Bern et al, determinants of VL transmission in the ISC included living in mud houses, proximity to prior cases (in the same or nearby household), presence of vegetation and standing water surrounding the house, sleeping on the floor or outside, malnutrition, and a lack of bed net use[33].

The greatest concentration of cases in East Africa is reported in the northern focus, specifically between the eastern Sudanese State of Gederaf and the bordering northern States of Ethiopia. Many of the epidemiological determinants of VL in East Africa are shared with the ISC, with poverty remaining the central overarching factor. Notable determinants include living in rural settings near sandfly breeding and resting sites (living near termite mounds in the southern focus, sleeping under *Acacia* trees in the northern focus), living in proximity to other VL infected (or recently infected) people, and malnutrition[34].

VL/HIV co-infection has reshaped VL epidemiology in many endemic regions, and remains the most important risk factor for asymptomatic infection, disease progression, and poor treatment outcomes. In the mid-1980s in Spain and other southern European countries, VL shifted from a rare childhood disease to one predominantly affecting HIV-positive adults[35, 36]. VL/HIV co-infection rates are currently increasing in Brazil (now reported in 20% of new cases[37]) and India (> 10% of VL episodes in 2023 and 2024[38]). In northern Ethiopia, VL/HIV co-infection presents a significant challenge to elimination, affecting between 20 and 50% of cases predominantly in male migrant workers[39].

Host immunity — particularly cell-mediated immunity — is a major determinant of disease onset and treatment outcomes, and thought to be a key driver of several devastating epidemics associated with war and natural disasters. A striking example is Sudan in the 1980s-2000s, where population displacement driven by conflict and

⁴Despite sporadic VL cases reported from Pakistan, Bhutan and Sri Lanka, in this thesis ISC refers to India, Nepal and Bangladesh.

famine brought immune-naïve populations into endemic areas, and spread infection into previously unaffected areas. Combined with severe malnutrition and the collapse of health infrastructure, the resulting mortality was catastrophic[13, 40, 41].

The geographic and demographic heterogeneity of VL reflects fundamental differences in host-parasite interactions and immune responses. To understand how these epidemiological patterns arise and why certain groups experience more severe disease, it is necessary to examine the pathophysiology of human *Leishmania* infection.

2.2.2 Pathophysiology

Parasites of the genus *Leishmania* belong to the family *Trypanosomatidae* (class *Kinetoplastea*, phylum *Euglenozoa*) and share many features with two other human pathogens of the same family; *Trypanosoma brucei* — the causative agent of human African Trypanosomiasis (African sleeping sickness), and *Trypanosoma cruzi* — the agent of Chagas disease in the Americas. These vector-borne NTDs are all single-celled protozoa characterized by a flagellum and a kinetoplast — a dense network of mitochondrial DNA giving them a distinctive microscopic appearance.

Life Cycle

In the sandfly gut, *Leishmania* parasites exist in their extracellular flagellated promastigote forms. During a blood meal, the parasite is regurgitated into the bite wound and rapidly phagocytosed by macrophages. Intracellularly, in the human host, the protozoa lose their flagellum and transform into amastigotes: oval bodies measuring $\sim 2\text{--}6\mu\text{m}$ in length with their characteristic nucleus and kinetoplast clearly visible on nucleic acid staining. Intracellular amastigotes multiply through binary fission, eventually leading to cell lysis and subsequent uptake by neighbouring macrophages to perpetuate infection. In viscerotropic disease caused by *L. donovani* complex, this cycle drives widespread dissemination with macrophage proliferation and granuloma formation across the mononuclear phagocyte system, including in the spleen, liver, bone marrow, and lymph nodes. The cycle completes when sandflies ingest parasitised macrophages during a subsequent blood meal, initiating parasite development within the vector (Figure 2.3).

Immune Evasion

Remarkably, intracellular amastigotes have evolved to survive and replicate within the hostile phagolysosomal environment of the macrophage. Although the full immunological mechanisms remain incompletely understood and beyond the scope

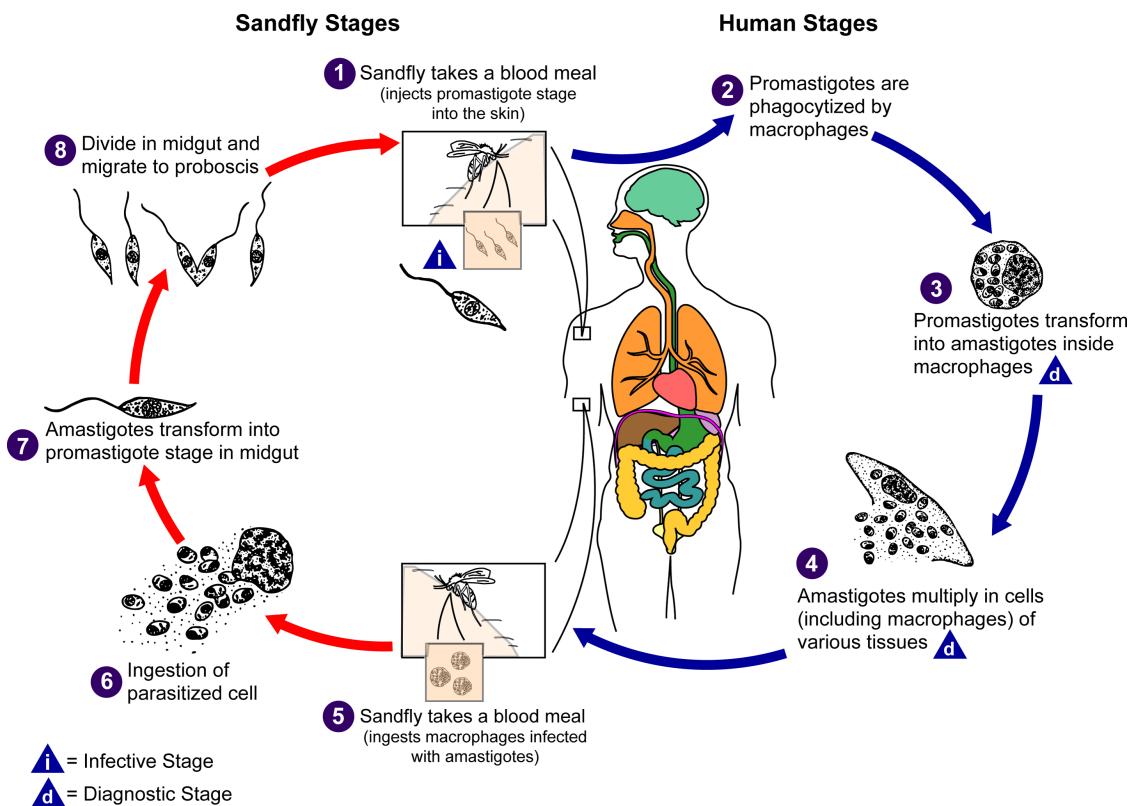


Figure 2.3: Life cycle of the *Leishmania* parasite consists of a vector (sandfly) stage and reservoir (animal) stage. While vector transmission is by far the most common route, it is not essential — direct human-to-human transmission has been reported via needle sharing, organ transplantation, blood transfusion, and mother-to-child transmission. Illustration credit: Centers for Disease Control and Prevention/Alexander J. da Silva/Melanie Moser.

of this chapter, key strategies include the presence of a lipophosphoglycan surface coat that protects promastigotes from complement-mediated lysis, and suppression of macrophage activation by inducing functional impairment of cytotoxic T-cell responses. Despite the presence of a pronounced humoral response, with a polyclonal hypergammaglobulinaemia typically seen, the resulting antibodies are predominantly non-neutralising and may even play a role in maintaining the immunosuppressive milieu required for amastigote survival[42, 43]. As the disease progresses, so too does the host's immunosuppressive state, driven by rising levels of immunoregulatory cytokines (especially IL-10 and TGF β) and the destructive remodelling of lymphoid architecture[43, 44].

2.2.3 Clinical Features

The clinical spectrum of *Leishmania* infection is remarkably broad, with asymptomatic carriers outnumbering symptomatic cases by a factor of 5–200 in endemic

settings[3, 31, 45]. If symptomatic, disease onset is typically insidious, following an incubation period ranging from several weeks, to months, and occasionally years[46].

Classic symptoms result from a persistent systemic inflammatory response syndrome with infiltration of the mononuclear phagocyte system, resulting in intermittent fever, weight loss, splenomegaly, often with accompanying hepatomegaly and, in parts of East Africa, lymphadenopathy. The spleen is usually non-tender on palpation, but massive splenomegaly ± hepatomegaly can result in severe abdominal pain. The greyish skin hyperpigmentation ('kala-azar' — literally 'black disease' or 'black fever' in Hindi), described in historical accounts of chronic disease in the ISC before effective therapy became available, is now uncommon[2, 7].

Characteristic laboratory abnormalities include raised inflammatory markers (CRP, ESR), pancytopenia, a polyclonal hypergammaglobulinaemia, and hypoalbuminaemia. Anaemia is often severe, normocytic, and normochromic, and likely multifactorial in origin, with splenic sequestration and anaemia of chronic disease thought to play important roles[47, 48]. Haemoglobin levels at presentation are frequently between 6–10 g/dL[47], resulting in pallor and contributing to weakness.

Without treatment, disease severity correlates with increasing parasite burden[49, 50]. Patients may develop vomiting, diarrhoea, cough, and dyspnoea, reflecting a combination of direct mucosal invasion, and opportunistic infections arising from the host's immunosuppressed state. The presence of oedema, jaundice, severe co-infection, and bleeding with disseminated intravascular coagulation indicates a poor prognosis[51].

2.2.4 Diagnosis

The gold standard for diagnosing VL combines compatible clinical features with direct visualisation of *Leishmania* amastigotes in tissue aspirates obtained from the spleen, bone marrow, or lymph node. Splenic aspiration offers the highest sensitivity (93–99%) but carries a small risk of fatal haemorrhage. Bone marrow and lymph node aspirates are safer but considered less sensitive (~50–80%)[2, 3]. Parasite density is usually expressed using a logarithmic semiquantitative grading system (0–6+) originally described by Chulay and Bryceson (Table 2.1)[52].

Culture of tissue aspirates can improve diagnostic sensitivity but is limited by cost, technical complexity, and a slow turnaround of up to four weeks. Molecular assays have been developed but remain infrequently available outside research settings and regions of high endemicity. Notably, peripheral blood quantitative polymerase chain reaction (qPCR) assays amplifying kinetoplast DNA (kDNA) have shown high sensitivities in patients with both VL and VL/HIV co-infection[53–55],

Table 2.1: Logarithmic grading system for parasite load on microscopy of tissue aspirates described by Chulay and Bryceson in the early 1980s[52]. Smears are stained with Giemsa and examined using a $10\times$ eyepiece and $100\times$ objective lens.

Grade	Average parasite density
0	0 amastigotes/1,000 fields
1+	1–10 amastigotes/1,000 fields
2+	1–10 amastigotes/100 fields
3+	1–10 amastigotes/10 fields
4+	1–10 amastigotes/field
5+	10–100 amastigotes/field
6+	> 100 amastigotes/field

obviating the need for invasive sampling where available. Urine antigen detection has demonstrated specificity but suffers from low diagnostic sensitivity[56, 57].

A wide gamut of serological (antibody-detecting) tests is available for the diagnosis of VL, including enzyme-linked immunosorbent assays (ELISA), indirect fluorescent antibody (IFA) assays, immunoblots, and rapid diagnostic tests (RDTs)[2]. The rK39 immunochromatographic test (ICT) is the most widely used RDT, providing a binary result within 10–20 minutes from a finger-prick blood sample. In the ISC, rK39 RDTs demonstrate excellent sensitivity (97.0%, 95% CI 90.0–99.5%)[58] and have served as the first-line diagnostic test since the mid-2000s. In East Africa, however, sensitivity is lower (85.3%, 95% CI 74.5–93.2%), and a second serological test — the direct agglutination test (DAT) — is recommended to confirm negative rK39 results[58]. Key limitations of serological tests include their inability to distinguish between active and past infection, with antibodies persisting for months to years following successful treatment and their reduced sensitivity in patients with severe immunosuppression, particularly those with VL/HIV co-infection[3].

2.2.5 Treatment

Intramuscular (IM) pentavalent antimonials have been the workhorse of VL treatment since the 1940s⁵, and despite their relatively toxic adverse effects (pancreatitis, cardiac arrhythmias, hepatitis), they remain first-line therapy in both East Africa (sodium stibogluconate [SSG], as part of combination therapy) and in Brazil (as meglumine antimoniate [MA])[59, 60]. From the early 1980s in Bihar, India, treatment failure rates exceeding 50% were observed with SSG despite dose escalation

⁵Previously, since the 1910s, tartar emetic was used — a toxic trivalent antimonial compound also used as an emetic.

(10→20 mg/kg/day) and duration (6→40 days). Blame was attributed to poor stewardship, with subtherapeutic dosing practices driving resistance[61]. As efficacy declined, second-line agents were used — initially IM pentamidine (diabetogenic, and limited by severe toxicity), and later intravenous (IV) amphotericin B deoxycholate (ABD), effective but constrained by infusion reactions, nephrotoxicity, and the need for prolonged hospitalisation.

Over the past two decades, the introduction of three new agents — oral miltefosine, IV liposomal amphotericin B (initially registered in 1997 as AmBisome®; Gilead Sciences), and IM paromomycin — has transformed the previously limited and often toxic treatment arsenal[62].

Miltefosine (MF), a repurposed anticancer drug, is the only effective oral agent for VL. Introduced as the first-line treatment in the ISC in 2005, MF played a pivotal role in the early KAE[63–65]. Its use has since declined. Concerns about teratogenicity, gastrointestinal toxicity, and decreasing efficacy[66, 67] led to its replacement in the KAE by single dose liposomal amphotericin B (LAMB) (AmBisome®; Gilead Sciences) in 2014–15[68]. Outside the ISC, MF performs poorly as monotherapy in East Africa and Brazil[62, 69].

Supported by a WHO–Gilead donation programme, a single intravenous IV dose of 10 mg/kg LAMB remains the first-line regimen in the ISC owing to its efficacy and safety[70]. At higher cumulative doses of 21–30 mg/kg given over 5–10 days, LAMB is also used as first-line therapy in Europe and as a second-line option in Brazil and East Africa, particularly in patients intolerant to pentavalent antimonials (severe disease, extremes of age, pregnancy, renal or hepatic dysfunction, relapse, VL/HIV co-infection).

Paromomycin (PM), an injectable aminoglycoside, is effective both as monotherapy and in combination regimens in the ISC[71]. Despite its low cost and favourable safety profile, uptake in the ISC has been limited. In East Africa, by contrast, SSG/PM combination therapy has been first-line since 2010[59, 72], with evidence also supporting its use alongside MF in both East Africa and the ISC[73, 74].

2.2.6 Elimination Efforts

In areas of anthroponotic transmission, control of VL centres on two interdependent strategies: (i) reduction of the human infection reservoir through early diagnosis and treatment, and (ii) interruption of transmission events through vector control measures. Effective implementation of these approaches requires evidence-based decision-making and sustained political will.

Guided by these principles, in 2005 the governments of Nepal, Bangladesh, and India, with WHO support, signed a Memorandum of Understanding to reduce VL incidence to fewer than 1 case per 10,000 population at district or sub-district levels by 2015. This deadline was later extended and is now embedded within the 2021–2030 WHO NTD roadmap[65, 75–78]. The initial success of the KAEP was attributed to the introduction of highly sensitive rK39 RDTs, oral MF, and subsequently single-dose LAMB. These advances were complemented by active case detection strengthened through community engagement and public awareness, including the use of frontline community health workers such as Accredited Social Health Activists (ASHAs) in India[79]. Vector control measures, including indoor residual spraying (IRS) and insecticide-treated nets (ITN), further supported transmission reduction[80]. In October 2023, Bangladesh became the first country to eliminate VL as a public health problem[81]. Nepal and India are currently working to sustain their targets for three consecutive years to achieve elimination status[82].

As the ISC enters the consolidation and maintenance phases of the KAEP, concerns regarding sustainability are increasingly voiced[68]. Waning political momentum and financial support threaten the viability of costly IRS and active case detection. Moreover, as the proportion of immunologically naïve individuals increases, so too does the risk of future outbreaks[83].

Buoyed by the KAEP experience, calls for VL elimination in East Africa[84] culminated in the 2023 *Nairobi Declaration*, endorsed by the health ministries of nine endemic East African countries⁶ and aiming to reduce VL incidence by 90% by 2030. Similar to the KAEP, a phased approach has been proposed — consisting of planning, attack, consolidation, and maintenance phases — with emphasis placed on equitable access to diagnosis and treatment, integrated vector management adapted to diverse ecological settings, and strong cross-border collaboration[85, 86].

However, in contrast to the KAEP, the East Africa initiative faces greater operational complexity. Challenges include heterogeneous transmission ecologies, weaker health systems, reliance on multi-dose treatments, limited access to sensitive diagnostic tools, and perhaps more importantly, political instability and the consequences of population displacement due to civil wars and famine[86].

Early diagnosis and treatment of all human infection reservoirs remains a cornerstone of any successful elimination campaign where transmission is largely anthropontic. The next section examines the burden, mechanisms, and determinants of VL relapse — a reservoir of particular concern and of critical relevance to elimination.

⁶Chad, Djibouti, Eritrea, Ethiopia, Kenya, Somalia, South Sudan, Sudan and Uganda

2.3 Relapse

This section is guided by a systematic review of the literature and presented as a narrative review (see Box 2.2). Search terms were constructed to identify the burden, timing, mechanisms, and determinants of VL relapse, with a focus on immunocompetent patients in the ISC and East Africa. Further details are available in Appendix A.

VL relapse is defined as the reappearance of VL signs and symptoms following an initial treatment response[87, 88]. In both research and routine clinical settings, relapse is typically confirmed by direct visualisation of the parasite on a tissue aspirate smear[89].

Relapse can only occur once an initial treatment response is achieved — termed initial cure in clinical efficacy studies. In research settings, this requires both clinical improvement (e.g., defervescence, reduction in spleen size, weight gain, improvement of anaemia) and confirmation of parasite clearance by microscopy. This is referred to as *test-of-cure*, and usually occurs within a month of the end of treatment (EoT), although considerable heterogeneity exists in precisely how and when test-of-cure is assessed[90].

Box 2.2: Literature search

Literature search performed from database inception to 11th August 2025 in PubMed (below), Embase and Web of Science. Articles in English were reviewed. Full search details available in Appendix A.

```
("Leishmaniasis, Visceral"[Mesh] OR "visceral leishmaniasis" OR  
"leishmaniasis, visceral" OR "kala azar" OR "kala-azar") AND  
("Recurrence"[Mesh] OR relapse* OR recurrent OR recurrence OR  
recrudescence OR "treatment failure")
```

2.3.1 Burden of Relapse

Reported relapse rates among immunocompetent patients vary widely across studies, reflecting differences in host, treatment, and parasite factors[89]. As a ‘rule of thumb’, relapse rates with current first-line regimens in patients without significant immunosuppression range between 2.5% to 10%[89]. According to a meta-analysis of VL clinical efficacy studies by Chhajed et al, the overall proportion of HIV-negative patients relapsing in the ISC within 6 months of treatment was estimated at 3.5% (95% confidence interval [CI]: 2.8–4.5%) following first-line treatment with single dose LAMB[89]. Under pragmatic conditions in East Africa, slightly higher

6-month relapse rates of approximately 5% are seen with the first-line combination therapy of PM and SSG[91–93].

These estimates, however, understate the true relapse burden: (i) relapses occurring beyond 6-months of follow-up are missed, and (ii) clinical trial populations do not reflect the broader patient population seen in routine care. Chhajed et al also showed that, across 21 studies reporting both 6 and 12 month relapse rates, one third of all relapses by 12 months occurred during the second half of the 12 month period[89]. Similarly, several large studies from the ISC have shown comparable or even higher relapses counts occurring between 6–12 months than in the initial 6 month period, prompting calls to extend routine follow-up from 6 to 12 months[66, 94–97]. Furthermore, trials often exclude patients at increased risk of relapse, such as those at the extremes of age, with more severe forms of the disease, and with comorbidities. Trial patients are also managed under controlled conditions that incentivise adherence. These factors collectively bias relapse estimates downward compared with outcomes observed under routine conditions.

2.3.2 Relapse vs. Reinfection

Distinguishing relapse resulting from parasite recrudescence, from re-infection with a new strain, is important when assessing drug efficacy and relapse determinants. However, unlike malaria[98], VL lacks validated molecular targets for PCR-based confirmation. Nevertheless, accumulating evidence indicates that most relapses result from recrudescence rather than reinfection. For example, Rijal et al. performed kDNA fingerprinting in 8 pairs of bone marrow samples in HIV-negative patients prior to primary infection treatment and at the time of relapse. No evidence of reinfection was found, with 8 distinct kDNA fingerprints identified across the 16 samples that matched at the individual patient level[66]. Similar studies have been performed in VL/HIV co-infection patients in East Africa and Europe employing a variety of molecular techniques, and showing that while reinfection does occur, it is considerably less common than recrudescence[99–102].

2.3.3 Relapse Determinants

When selecting candidate predictors for inclusion in a prognostic model, it is important to draw on previously described predictors[103]. To provide important context for the subsequent chapters on relapse model development, this section summarises the determinants of relapse identified through the literature review.

For immunocompetent patients, the review aims to be exhaustive. Studies were excluded if they (i) did not adopt a longitudinal design in which index VL cases were

linked to subsequent relapse episodes in the same patients, (ii) included more than 5% VL/HIV co-infection without reporting relapse predictors separately by HIV status; or (iii) focused solely on treatment regimens and/or biomarkers (including molecular targets or cytokines) that are not routinely measured in clinical practice. Although not intended to be comprehensive, key studies conducted in patients with VL/HIV co-infection, as well as those reviewing biomarkers or treatments predictive of relapse, are also discussed.

Four systematic reviews describing predictors of relapse were identified: two including patients with and without VL/HIV co-infection[104, 105] and two focussing specifically on patients with VL/HIV co-infection [106, 107]. Where appropriate, references from these reviews inform the narrative synthesis.

It is important to recognise that whether a study identifies a significant association between a predictor and relapse does not necessarily correspond to whether or not the predictor is important. For example, (i) not all predictors are considered by all studies, (ii) many studies have small sample sizes and are therefore underpowered to detect certain associations, and (iii) a wide range of study designs, statistical tests, predictor transformations, and modelling strategies are employed, affecting what is considered ‘significant’. Furthermore, the significance of an association will also depend on the other variables adjusted for in the prediction model.

Immunocompetent Patients

A total of 11 studies reporting relapse determinants in immunocompetent patients were identified, and are presented in Table 2.2. Eight studies were conducted in the ISC, including India[94, 96, 97, 108, 109], Bangladesh[110, 111], and Nepal[66], two in East Africa, including South Sudan[112] and Kenya[113], and one in Georgia[114]. The median study size was 1,143 patients (range 115[66] to 8,537[96]). Two studies presented outcomes from the same trial[97, 108], although they are reported separately here due to differences in sample size and methodology. Follow-up ranged from 6 months to 4 years, and relapse proportions from 2.6%[110] to 20.8%[66]. Further details, including study design and methodologies, are provided in the [Supplementary Material](#).

Age, treated categorically (≥ 2 levels), was the most frequently reported relapse determinant. All 11 studies considered age as a candidate predictor, with eight demonstrating statistically significant ($p < 0.05$) or borderline significant ($0.05 \leq p < 0.1$) associations in unadjusted models[66, 94, 96, 97, 108–110, 114]. When adjusted models were considered, age remained significant in all but one study[94]. Two ISC studies that modelled ≥ 3 age categories showed a U-shaped relationship, with

increased relapse risk in young children (< 5 years) and older adults (≥ 40 or ≥ 45 years)[96, 110]. Other studies identified increased risk only in younger groups; < 12 or < 15 years in the ISC[66, 94, 97, 108] and < 1 year in Georgia[114]. Notably, the two East African studies did not identify age as a significant predictor[112, 113], although this may reflect limited power (17 relapses in Kennedy et al.[113]), or selection bias due to incomplete linking of relapse and index cases[112]. No study explicitly modelled age as a continuous variable.

Spleen size was the next most frequently identified determinant, considered in seven studies. Five reported significant associations in unadjusted analyses, and four in adjusted models[94, 96, 110, 112, 114]. Definitions and timing varied. Lucero et al. reported that larger spleens at discharge predicted relapse (OR 1.27, 95% CI 1.10–1.47 per cm below the costal margin). Sundar et al. found approximately twice the odds of relapse in patients with admission spleen size > 4 cm[94]. Gorski et al. identified increased relapse odds with larger spleens at discharge *and* admission, when measured with Hackett grade[112]. Kajaia et al. reported increased relapse risk with larger spleens at admission when measured by the Kandelaki splenometric method[114, 115]. Instead of absolute spleen size, Burza et al. modelled change in spleen size during admission: patients with a reduction ≤ 0.5 cm/day had 1.7-fold higher odds of relapse (95% CI 1.1–2.5) compared with those with greater reductions[96].

Symptom duration prior to treatment was considered in six studies, with four identifying significant associations in both unadjusted and adjusted models[96, 97, 108, 114]. In the ISC, Burza et al. reported that patients with ≤ 4 weeks of symptoms had higher relapse odds (1.6-fold vs 4–8 weeks; 2.3-fold vs > 8 weeks), with similar effects after adjusting for age, sex, and change in spleen size[96]. Goyal et al. found that ≤ 8 weeks of symptoms was associated with 3.3-fold (95% CI 1.3–8.4) higher relapse odds[97] and a 3.6-fold (95% CI 1.4–9.1) higher relapse hazard in time-to-event analysis[108]. In contrast, Kajaia et al. reported that ≥ 90 days of symptoms was associated with higher relapse risk in Georgia (OR 3.9; 95% CI 1.8–8.5) after adjusting for haemoglobin and age[114]. Kennedy et al. included symptom duration but did not identify a significant association, perhaps due to low statistical power[113].

Although sex was considered in all studies, only three ISC studies found significant associations with relapse[94, 96, 109]. Each reported an approximate doubling in relapse odds[94, 96] or rates[109], with similar effect sizes seen in adjusted analyses.

Admission haemoglobin (Hb) was included in eight studies[66, 94, 96, 97, 110, 113, 114] and was significant in three[110, 113, 114]. Lucero et al. observed strong associations between lower Hb and relapse in partially adjusted models, although these were not retained in the fully adjusted model (OR 1.40, 95% CI 1.09–1.72 and 1.45, 95% CI 1.15–1.85, per 10 g/L decrease, for admission and discharge Hb, respectively). Kajaia et al. reported a strong association, with relapse 12-fold (95%CI 4.1–34.8) more likely in patients with Hb < 60 g/L vs \geq 80 g/L, persisting after adjustment for age and symptom duration.

Markers of malnutrition were inconsistently defined across studies, and reporting was often unclear. Gorski et al. and Burza et al. used age-specific definitions combining BMI, BMI-for-age z-scores, and weight-for-height z-scores[96, 112]. Lucero et al. used a similar approach, combining weight-for-height z-scores and BMI[110]. Goyal et al. and Kennedy et al. described ‘severe wasting’ and ‘malnutrition’, respectively, without further definitions[97, 113]. Rijal et al. and Sundar et al. relied on single indicators (BMI and weight, respectively)[66, 94]. Only Sundar et al. observed a significant association, with patients \leq 30 kg at higher risk, although the causal interpretation is likely confounded by age, given that the study recruited participants of all ages[94].

Table 2.2: Summary of studies presenting significant predictors of VL relapse in immunocompetent patients. Excluding: studies only looking at treatment regimen or biomarkers. BPKIHS: B.P. Koirala Institute of Health Sciences; ISC: Indian subcontinent; KAMRC: Kala-azar Medical Research Centre; MA: meglumine antimoniate; MSF: Médecins Sans Frontières; Hb: haemoglobin; LAMB: liposomal amphotericin B; MF: miltefosine; PM: paromomycin; RBC: red blood cell; SDA: single dose liposomal amphotericin B 10mg/kg; SSG: sodium stibogluconate; VL: visceral leishmaniasis; wks: weeks; yrs: years.

Study	Location	Treatment	n (relapse) %	Significant predictors — unadjusted model ¹	Significant predictors – adjusted model ²
ISC					
Burza 2014[96]	MSF clinics, Bihar, India	20 mg/kg LAMB (4 × 5 mg/kg)	8,537 (119) 1.4%	age < 5****, age ≥ 45** vs. 15–30 yrs; male sex***; other backwards caste* vs. general category; symptom duration ³ > 8 wks***, > 4 to ≤ 8 wks** vs. ≤ 4 wks; spleen size change ≤ 0.5 cm/day vs. > 0.5 cm/day	age < 5****, age ≥ 45** vs. 15–30 yrs; male sex***; symptom duration ³ > 8 wks**, > 4 to ≤ 8 wks** vs. ≤ 4 wks; spleen size change ≤ 0.5 cm/day vs. > 0.5 cm/day
Goyal 2019[97]	Public health facilities, Bihar, India	SDA; LAMB+MF; MF+PM	1,353 (75) 5.5%	age 2–12 yrs** vs. > 12 yrs; symptom duration ≤ 8 wks*** vs. > 8 wks; treatment***	age 2–12 yrs [†] vs. > 12 yrs; symptom duration ≤ 8 wks [†] vs. > 8 wks; treatment MF+PM [†] vs. SDA
Goyal 2020[108]	[as above]	[as above]	1,750 (79) 4.5%	age 2–12 yrs ^{††} vs. > 12 yrs; treatment with MF+PM ^{††}	age 2–12 yrs*** vs > 12 yrs; symptom duration ≤ 8 wks*** vs. > 8 wks; treatment MF+PM**** vs. SDA
Lucero 2015[110]	MSF clinic, Fulbaria, Bangladesh	15 mg/kg LAMB (3 × 5 mg/kg)	1,521 (39) 2.6%	age < 5 yrs***, age ≥ 40 yrs** vs. 18–39 yrs; lower admission Hb***; lower discharge Hb***; larger discharge spleen size***	age < 5 yrs***, age ≥ 40 yrs** vs. 18–39 yrs; larger discharge spleen size**
Mondal 2019[28]	Hospital in Mymensingh, Bangladesh.	8 different regimens	984 (69) 7.0%	treatments compared only: MD-LAMB*, MF**, LAMB+PM**, LAMB+MF***, SDA****, PM****, MF+PM**** vs. SSG	treatment: MD-LAMB*, MF**, LAMB+PM**, SDA***, LAMB+MF****, PM****, MF+PM**** vs. SSG
Rijal 2013[66]	BPKIHS, Eastern Nepal	MF	115 (24) 20.9%	age ≤ 12 yrs** vs. > 12 yrs	[not performed]
Sundar 2019[94]	KAMRC, Bihar, India	SDA	1,143 (66) 5.8%	age ≤ 15 yrs* vs. > 15 yrs; male sex**; weight ≤ 30 kg** vs. > 30 kg; spleen size > 4 cm*** vs. ≤ 4 cm	male sex***; weight ≤ 30 kg** vs. > 30 kg; spleen size > 4 cm*** vs. ≤ 4 cm

continued on next page

Table 2.2: continued

Study	Location	Treatment	n (relapse) %	Significant predictors — unadjusted model ¹	Significant predictors – adjusted model ²
East Africa					
Gorski 2010[112]	MSF clinics, South Sudan	SSG, SSG/PM, LAMB	8,090 (166) ⁴	treatment centre: Lankien **** vs other; admission spleen size ^{5***} ; discharge spleen size ^{***} ; treatment ^{****} ; lower % gain in body weight [*]	admission spleen size ^{5†, 2†, ≥ 3†} vs 0; discharge spleen size ^{2†, ≥ 3†} vs 0; treatment SSG/PM [†] vs SSG
Europe					
Kajaia 2011[114]	Institute of Parasitology and Tropical Medicine, Tbilisi, Georgia	MA	300 (21) 7.0%	age ^{**} : higher risk in < 1 yrs; symptom duration ^{***} : higher risk with longer duration; Hb ^{****} : higher risk with Hb < 60 g/L; RBC [*] : higher risk < $2.0 \times 10^{12}/\text{L}$; lymphocytes ^{**} : higher risk with $\geq 80\%$; admission spleen size ^{**} : higher risk with larger spleens	age < 1 yr ^{**} vs. > 1 yr; symptom duration ≥ 90 days ^{***} vs < 90 days; Hb < 60 g/L ^{****} vs. Hb ≥ 80 g/L

¹ Significant predictors in Lucero et al. are in fact adjusted for sex and age. Kajaia et al present the overall p-value across all predictor categories and 95% CIs for each category. Overall p-values are presented, and the trend commented on.

² 3 studies adjusted the multivariable model for other non-significant variables, including Lucero et al. (sex, admission and discharge haemoglobin, admission spleen size), Mondal et al. (sex, age, treatment, and various others), Gorski et al. (age, sex, calendar year, treatment centre)

³ Shorter symptom duration associated with higher relapse risk in Burza 2014.

⁴ 166 relapses cases could be matched with primary case records, and does not reflect total relapse number.

⁵ Spleen size recorded as Hackett grade in Gorski 2010.

* $0.05 \leq p < 0.1$.

** $0.01 \leq p < 0.05$;

*** $0.001 \leq p < 0.01$

**** $p < 0.001$.

† $p < 0.05$ (inferred as confidence interval does not cover null).

†† Significance level not stated.

VL/HIV Co-infection

Relapse rates in patients with VL/HIV co-infection often exceed 20–50%[106, 116, 117], highlighting the importance of host immunity in achieving lasting cure.

Two systematic reviews were identified that summarised relapse determinants in this immunosuppressed population, drawing on 18 studies (Cota et al., 2011)[106] and 15 studies (Alemayehu et al., 2016)[107], respectively. Significant predictors of relapse included low baseline CD4⁺ T-cell counts (particularly < 100 cells/ μ L), previous history of VL relapse, lack of improvement in CD4⁺ counts during follow-up, absence of secondary prophylaxis, and the late initiation of highly active antiretroviral therapy (HAART) initiation.

Since the publication of these reviews, several important studies have shed further light on relapse determinants in VL/HIV co-infected patients[116, 118, 119]. In 2017, Abongomera et al. reported outcomes from a retrospective study of 146 co-infected patients treated at an MSF-supported clinic in Northwest Ethiopia between 2008 and 2013, in which 44 (30.1%) relapsed during follow-up. In addition to the timing of HAART initiation, further predictors included high baseline tissue parasite load (adjusted hazard ratio [aHR] 6.63, 95% CI 2.64–16.63, 6+ vs \leq 6+) and the presence of splenomegaly on admission (borderline significant in unadjusted analysis)[118].

More recently, Costa et al. (2023) published a prospective study of 169 co-infected patients from Maranhão, Brazil (2013–2020), in which 70 (41.1%) relapsed during 12 months of follow-up. Baseline splenomegaly, lymphadenopathy, previous VL relapse, HAART regimen, HAART duration, elevated creatinine, and elevated urea were all statistically significant predictors of relapse in unadjusted analyses[119].

In contrast to findings in immunocompetent patients, only one study identified age as a statistically significant predictor of relapse. Burza et al. demonstrated an approximate doubling of relapse hazard among patients \geq 40 years, using routinely collected data from MSF treatment centres in Bihar, India (2007–2012)[117].

Treatments

After VL/HIV co-infection, treatment is arguably the second most important predictor of relapse. This is evident across numerous studies, where the most common reason for failing to achieve the primary outcome of definite cure — defined as initial cure and relapse-free survival for at least 6 months — is relapse[62, 72, 73, 94, 108]. Increases in relapse rates have been suggested as a sensitive indicator of declining treatment efficacy and the emergence of drug resistance, often preceding overt increases in primary treatment failure[120]. Summarising

outcomes from all VL efficacy studies is beyond the scope of this thesis, although several observations merit discussion.

Firstly, as summarised by Chhajed et al., although relapse risk varies considerably across regimens, combination therapies are generally associated with lower relapse risk than monotherapies[89]. After adjusting for age and symptom duration, Goyal et al. showed that in India, treatment with paromomycin and miltefosine combination therapy was associated with significantly reduced odds of relapse compared with single dose LAMB 10 mg/kg (aOR 0.21, 95% CI 0.08–0.55). Similarly, in a large cohort study from Bangladesh, Mondal et al. found that, with the exception of SSG and multiple-dose LAMB, combination regimens were associated with lower relapse rates compared with paromomycin and miltefosine monotherapies[111]. A possible exception to this pattern is SSG/PM combination therapy. Using programmatic MSF data from South Sudan, 1999–2007, Gorski et al. demonstrated an approximate doubling of relapse odds in the SSG/PM group compared to SSG monotherapy (OR 2.08, 95% CI 1.21–3.58)[112].

Secondly, drug dose and treatment duration are also important predictors of treatment failure and relapse. Chhajed et al. showed this at the aggregate level by performing a meta-regression of relapse risk in the ISC against the dose of single dose LAMB, demonstrating that higher LAMB doses were associated with lower relapse risk (OR 0.81 per 1 mg/kg increase, 95% CI 0.72–0.91)[89]. Similar relationships between dose and overall treatment failure (lack of initial cure or subsequent relapse) have been demonstrated under trial conditions for both LAMB[121] and paromomycin[122, 123]. In East Africa, Dorlo et al. used a paediatric pharmacokinetic–pharmacodynamic (PK–PD) model to show an inverse relationship between miltefosine exposure and relapse hazard, with similar findings reported in the ISC[124]. These results supported the development of allometric dosing strategies for paediatric miltefosine[125, 126].

Lastly, both the intrinsic infectivity of the parasite strain and the development of drug resistance contribute to increasing rates of initial treatment failure and, when initial cure is achieved, subsequent relapse. A well-known example is the emergence of antimony resistance in Bihar, India, where treatment failure and relapse rates following treatment with SSG increased from < 5% to > 50% during the 1980s and 1990s, despite incremental escalation in SSG dose and duration[61, 127]. These failures have been attributed to both antimony resistance[128], and also to the selective survival and transmission of more infectious parasite strains[129]. Similar concerns arose following the widespread introduction of miltefosine in the ISC in the late 2000s, when increasing relapse rates were observed[66, 67, 130]. Interestingly,

comparisons of relapsing versus non-relapsing strains revealed no evidence of drug resistance, although relapsing strains were found to be more infectious[131]. More recently, Naylor-Leyland et al. reported increasing relapse rates between 2001 and 2018 that could not be explained by patient-level factors, prompting concerns about declining SSG/PM efficacy potentially due to emerging resistance[132].

Biomarkers

Since tissue aspirate sampling is invasive, substantial effort has been directed toward developing biomarkers that can serve both as a test-of-cure following treatment and as a diagnostic test of relapse[88]. In fact, a 2015 systematic review identified 53 biomarkers from 170 studies with potential to be used for monitoring post-treatment outcomes in Leishmaniasis[133]. Biomarkers can be broadly categorised into two categories: (1) direct markers of parasite burden, including DNA and antigen detection; and (2) indirect markers of host immunity, such as antibodies, cytokines and acute-phase proteins.

A large portion of the VL biomarker literature focuses on the direct detection of *Leishmania* DNA/RNA in both immunosuppressed[134–139] and immunocompetent patients[55, 140–143]. In particular, real-time quantitative PCR (qPCR) targeting kDNA, performed on the buffy coat of peripheral blood, has been confirmed as a strong predictor of relapse[55, 140, 142, 143], although routine use is limited by the need for specialised laboratory expertise and high costs. In a landmark 2021 study, Verrest et al. used qPCR measurements from 177 immunocompetent patients enrolled in DND*i* trials in East Africa⁷ and demonstrated that qPCR levels measured between treatment initiation and day 56 significantly predicted relapse, with areas under the receiver operating curve (AUC) of 0.71, 0.74, and 0.92 on days 14, 28, and 56, respectively[143]. Using the same qPCR data, Verrest et al. (2024) developed a semi-mechanistic population PK–PD model describing parasite replication, drug action, and post-treatment parasite clearance. The model successfully predicted relapse based on modelled day 28 and 56 parasite loads and provided the first direct evidence that relapse risk depends not only on initial parasite clearance but also on subsequent host immune responses[55]. Interestingly, inclusion of haematological and biochemical parameters,⁸ however, did not account for variation in post-treatment qPCR values.

⁷LEAP 0714 (30 children treated with allometric miltefosine), LEAP 0208 (151 patients receiving miltefosine plus LAMB combination regimens), FEXI VL 001 (14 patients receiving fexinidazole)

⁸including Hb, white blood cells (WBC), platelets, and creatinine

Beyond molecular testing, several studies have shown an association between relapse and the degree of *Leishmania* antigenuria in patients with VL/HIV co-infection[56, 57]. In Ethiopian patients, higher levels of urinary antigen at test-of-cure were strongly associated with relapse over the following 12 months (HR 9.8, 95% CI 1.8–82.1, comparing 1–10 parasites/10 fields with no parasites detected using the KATex assay).

Post-treatment IgG subclass titres have also been linked to relapse in both immunocompetent and immunosuppressed individuals[144–147]. However, because of the long half-life of immunoglobulins, their ability to predict relapse at the time of initial cure assessment remains uncertain. In asymptomatic patients, however, serological studies have shown that individuals with higher antibody titres are more likely to progress to symptomatic disease[148, 149].

A variety of non-serological and indirect biomarker panels have been evaluated for assessing disease activity during asymptomatic infection, active disease, and post-treatment recovery[150–155]. In East African patients, Kip et al. reported that the change in neopterin, a marker of macrophage activation, between EoT and day 60 of follow-up was predictive of relapse, with an AUC of 0.84[154]. In a subsequent study of 34 Ethiopian VL/HIV co-infected patients, half of whom relapsed, Takele et al. showed that relapse was associated with (i) failure to restore antigen-specific IFN γ production, (ii) persistently low CD4 $^{+}$ T-cell counts, and (iii) high T-cell expression of PD1. These results underscore the importance of effective cell-mediated immunity in achieving sustained cure.

Summary of Relapse Determinants

A substantial and rapidly growing evidence base has been identified describing the determinants of VL relapse, although methodological heterogeneity complicates synthesis.

Among immunocompetent patients — the focus of this thesis — the most consistently identified predictors are clear. Extremes of age (young children and older adults), larger spleen size, whether measured at admission or at EoT, male sex, and the severity of anaemia (at admission and EoT) are repeatedly associated with higher relapse risk. In contrast, evidence for malnutrition is weak, perhaps in part due to the varying definitions, and nuances in how measures are adjusted for age. Symptom duration shows conflicting associations. Shorter pre-treatment symptom intervals were strongly linked to higher relapse risk in two ISC studies, yet an opposite association was reported in the study from Georgia — highlighting potential contextual or methodological modifiers.

In patients with VL/HIV co-infection, predictors align more with immunological plausibility: markers of profound immunosuppression (notably low baseline CD4⁺ counts) and indicators of inadequate immune reconstitution (including timing and duration of HAART, absence of secondary prophylaxis) were the dominant determinants of relapse identified.

As anticipated, treatment-related factors are also highly influential. Treatment regimen, dose, and duration modify both initial cure and subsequent relapse risk, and therefore any prognostic model should be accounting for treatment variables where possible.

Finally, biomarker research (particularly peripheral blood qPCR) shows promising predictive performance, but current approaches are constrained by cost, laboratory requirements and limited routine availability, restricting their immediate applicability in many endemic settings.

2.4 Summary

VL remains a highly neglected tropical disease, disproportionately affecting the poorest and most marginalised communities. Despite substantial gains over the last two decades — most notably in the ISC, where coordinated elimination efforts have driven steep reductions in incidence — VL continues to impose a considerable clinical and economic burden. Sustaining these achievements is challenging, and translating success to high-burden regions such as East Africa remains an urgent global health priority[77].

Relapse is a critical, yet often understated, component of this challenge, with consequences both at the individual patient level and for public health more broadly.

For the patient, relapse is associated with heightened morbidity and mortality, exposure to prolonged and often toxic second-line therapies, and considerable direct and indirect financial costs for already vulnerable households.

From a public health and elimination perspective, relapse carries further implications. In anthroponotic transmission settings, early identification and effective treatment of relapse cases are essential to interrupt transmission. Moreover, parasite strains isolated from relapse cases can harbour resistance mutations and demonstrate greater fitness, underscoring their potential role as a consequential parasite reservoir.

Given the clinical, economic, and epidemiological significance of relapse, there is a clear and globally recognised need to improve the identification of individuals at greatest risk[86, 88]. Motivated by this, the overarching aim of this thesis is to determine whether clinical trial data contained within the IDDO VL data platform

can be leveraged to support the risk stratification of patients who are likely to progress to relapse following initial cure. In particular, this thesis explores whether such data can inform the development of a robust prognostic model for relapse.

If a prognostic model for VL relapse already exists, data from the IDDO VL data platform could be used to validate and, where necessary, update the model[103]. Therefore, to establish the current evidence base, the next chapter presents a systematic review of prognostic models developed for VL, encompassing both relapse-specific models and those predicting broader clinical outcomes.

3

Prognostic Models Predicting Clinical Outcomes in Patients Diagnosed with Visceral Leishmaniasis: A Systematic Review

Context

This systematic review was initially submitted for publication to *BMJ Public Health* in early 2023. Following editorial and peer review feedback received in December 2025, the literature search was updated and substantial revisions were made. The version presented in this chapter is adapted from the updated manuscript.

A protocol for this systematic review has been published^[156] and prospectively registered (PROSPERO registration number: CRD42023417226). Preliminary findings were disseminated through oral presentations at MSF Scientific Day (London, 2024) and the ASTMH 2024 Annual Meeting.

Authorship and Acknowledgements

While the manuscript was prepared by myself, the following co-authors contributed to this work (listed in publication order):

Forhad Chowdhury;¹ Shermarke Hassan;¹ Elinor Harriss;² Fabiana Alves;³

¹Infectious Diseases Data Observatory, Nuffield Department of Medicine, University of Oxford; University of Oxford Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford.

²University of Oxford, Bodleian Health Care Libraries

³Drugs for Neglected Diseases Initiative

Ahmed Musa;⁴ Prabin Dahal;¹ Kasia Stepniewska;¹ Philippe Guérin.¹

PG, JW, KS, and PD conceptualised the study. The final manuscript was written by JW, with review and input from all co-authors. Disease-specific expertise was provided by FA, AM, and PG. The search strategy was developed and executed by EH, with the literature search conducted by JW and FH. The data collection tool was designed by JW, and data extraction was performed by JW and SH.

In contrast to other chapters in this thesis, and to reflect co-author contributions, the plural first-person pronoun “we” is used throughout.

3.1 Introduction

In endemic settings, accurately identifying patients at high risk of adverse outcomes is paramount when prioritising limited resources, including admission, treatment selection, and follow-up intensity. Prognostic models — most commonly developed using multivariable regression techniques — estimate an individual patient’s probability of experiencing a future clinical event[157]. Often presented as simplified risk scores, such models abound in the medical literature and inform clinical decision-making and guideline development[158]. In infectious diseases alone, systematic reviews have identified over 600 prognostic models for COVID-19[159], 37 models for tuberculosis[160], and 27 models for malaria[161]. However, concerns have been raised on the methodological quality and reporting of prediction models. Biased models may overestimate performance, generate misleading predictions, and ultimately contribute to suboptimal or inequitable clinical decisions[158, 162].

Several prognostic models have been developed and implemented for patients with VL[163, 164]. In Brazil, national guidelines introduced in 2011 recommended the use of four related risk scores, based on combinations of clinical and/or laboratory factors, to guide hospital admission and the use of liposomal amphotericin B[164–166]. Similarly, since 2003, Médecins Sans Frontières (MSF) Holland has used simple VL risk scores in South Sudan to support clinical decision-making regarding liposomal amphotericin B therapy, broad-spectrum antibiotics, blood transfusions, and nutritional support[13, 112, 163]. Additional models exist[167, 168], although the full range of models, including their characteristics, comparative performance, and inherent biases, have yet to be systematically described.

We therefore conducted a systematic review to identify, summarise, and critically appraise prognostic models predicting future clinical outcomes in patients with VL. This review aims to support policymakers in evaluating the incorporation

⁴Institute of Endemic Diseases, University of Khartoum

of prognostic models into treatment guidelines, and to help clinicians assess the applicability of existing models to their own patient populations. In addition, researchers may use this review to identify evidence gaps and to determine whether available data are better suited to the development of new models, or to the external validation (evaluation) and/or updating of existing models[169].

A glossary of key terms relating to model development and evaluation is presented in Table 3.1.

Box 3.1: Key Messages of the Systematic Review

What is already known on this topic

- Visceral leishmaniasis (VL) is a neglected tropical disease associated with high mortality, predominantly affecting populations in resource-constrained settings.
- Identification of patients at high risk of adverse outcomes is critical for prioritising limited healthcare resources, including hospital admission, treatment selection, and follow-up care.
- Risk stratification of patients can be performed using prognostic models, however, the range of available models and their key methodological characteristics have not been systematically evaluated.

What this study adds

- Using existing reporting guidelines for systematic reviews of prediction model studies, this review provides the first comprehensive synthesis of prognostic models predicting future clinical outcomes in patients with VL.
- We identified 12 prognostic models, all of which predict mortality and were developed in Brazil or East Africa.
- All identified models and model evaluations were judged to be at high risk of bias; therefore model predictions and performance measures should be interpreted with caution.

How this study might affect research, practice or policy

- This review highlights key evidence gaps in the VL prediction model landscape and supports researchers in identifying candidate models for external validation or updating using local patient data.
- By summarising and appraising available models, this review enables clinicians and policymakers to assess the applicability of existing models to their own patient populations.
- By identifying common methodological limitations, this review encourages researchers to review contemporary guidance on the reporting of prediction model development and evaluation.

3.2 Methodology

We adhered to Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis: Checklist for Systematic Reviews and Meta-Analyses (TRIPOD-SRMA) when reporting this systematic review[170]. Data extraction was guided by the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS)[171] and the Prediction Model Risk of Bias Assessment Tool (PROBAST)[172, 173]. Risk of bias assessment is performed with PROBAST[172].

Eligibility Criteria

We followed a Population, Index model, Comparator model, Outcomes, Timing, Setting (PICOTS) approach to frame our review question and define our eligibility criteria[171, 174].

The population consisted of all human patients with a confirmed or suspected diagnosis of VL, as defined by the study authors. Index models included all prognostic models developed in patients with VL, including model development studies, external validation studies and/or model updating studies. No individual comparator model was defined, given the aim of the review was to summarise and critically appraise all identified model developments and evaluations. All clinical outcomes were considered that occur following the intended time of model use, with no upper limit on the prediction horizon. Timing of model use was either at the time of or following diagnosis. We imposed no restriction on the setting of model development or evaluation.

In accordance with best practice in prediction modelling research[172, 175, 176], we defined a prognostic model as a multivariable model (including two or more predictors) developed with the intention of predicting future outcomes at the individual patient level. Prediction model studies were distinguished from predictor finding or prognostic factor studies, where the aim is to investigate the effect of a single or group of factors on an outcome of interest[177]. We excluded unpublished studies (including conference abstracts, educational theses), studies that only report diagnostic prediction models, and animal studies.

Information Sources and Search Strategy

An information specialist (EH) created the search strategy to retrieve relevant records from the following databases: Ovid Embase; Ovid MEDLINE; the Web of Science Core Collection, SciELO and LILACS. No language restrictions were imposed. The databases were initially searched from database inception to 1 March 2023. Using the same strategy, the search was updated to include additional records from 1 March 2023 to 18 December 2025. The search strategy used text words and relevant indexing terms to retrieve studies describing eligible prognostic models. The Ingui search filter was augmented with an additional search string as described by Geersing et al[178, 179], and combined with VL-specific keywords. Google Scholar was used to identify any complementary grey literature (full search strategy presented in the [Supplementary Material](#)).

Study Selection

Deduplication and screening of references were performed in Covidence[180]. Screening was performed independently by two reviewers (JW, FC); initially at title and abstract level, and subsequently at full-text level. Where discordance existed, a third expert reviewer (PD) was consulted to make the final judgement. Subsequent forward and backward citation searching were performed to identify records missed by the initial search.

Data Collection

Study information was captured using a REDCap server hosted at the University of Oxford[181]. A data extraction form was created and piloted as per the CHAMPS checklist and PROBAST ([Supplementary Material](#))[171–173]. Two reviewers (JW, SH) independently extracted the study information. Where discordance remained after discussion, a final decision was made by a third expert reviewer (PD). Study authors were not contacted in the event of unclear or missing information.

Risk of Bias Assessment

Risk of bias was assessed using PROBAST[172, 173]. Two reviewers (JW, SH) independently assessed each model development (including updating) and external validation, by answering 20 signalling questions across four domains (participants, predictors, outcome, and analysis). Responses were used to judge the overall risk of bias as either ‘low’, ‘high’ or ‘unclear’. Any discordance was resolved through discussion.

As part of the bias assessment, we established whether the model predictors and their corresponding risk scores were consistent with the reported multivariable regression coefficients (PROBAST signalling question 4.9). Consistency was assessed by referring to the study's reported methodology, and expert guidance on the presentation of clinical risk scores[182].

Applicability

Applicability was not formally assessed given the broad remit of the research question. Instead, we summarise a broad range of information to facilitate comparison of each model's predictors, participants, and outcomes, with those of the intended target setting[173].

Synthesis of Results

Key characteristics of the identified models and their evaluations are summarised at the aggregate study and model levels, with complementary information presented in accompanying tables and [Supplementary Material](#). As all identified models predict mortality, a figure is included to facilitate comparison of candidate and final predictors across models.

A complementary narrative summary is also provided in the [Supplementary Material](#), where each study is described individually to provide additional contextual detail.

All identified models reported discrimination using the c-statistic. Accordingly, median and range values are presented for models evaluated in the development dataset (apparent performance) and in new patient data (external validation). Where c-statistics or other measures were not explicitly reported, no attempt was made to derive them from other performance measures. Owing to sparse and heterogeneous reporting across studies, other reported performance measures were not suitable for quantitative synthesis and are instead reported separately for individual models where available.

Meta-analysis was not performed due to substantial heterogeneity across models, including differences in predictors, populations, and outcome definitions, precluding meaningful pooling of c-statistics.

Patient and Public Involvement

No patients or members of the public were involved in this systematic review.

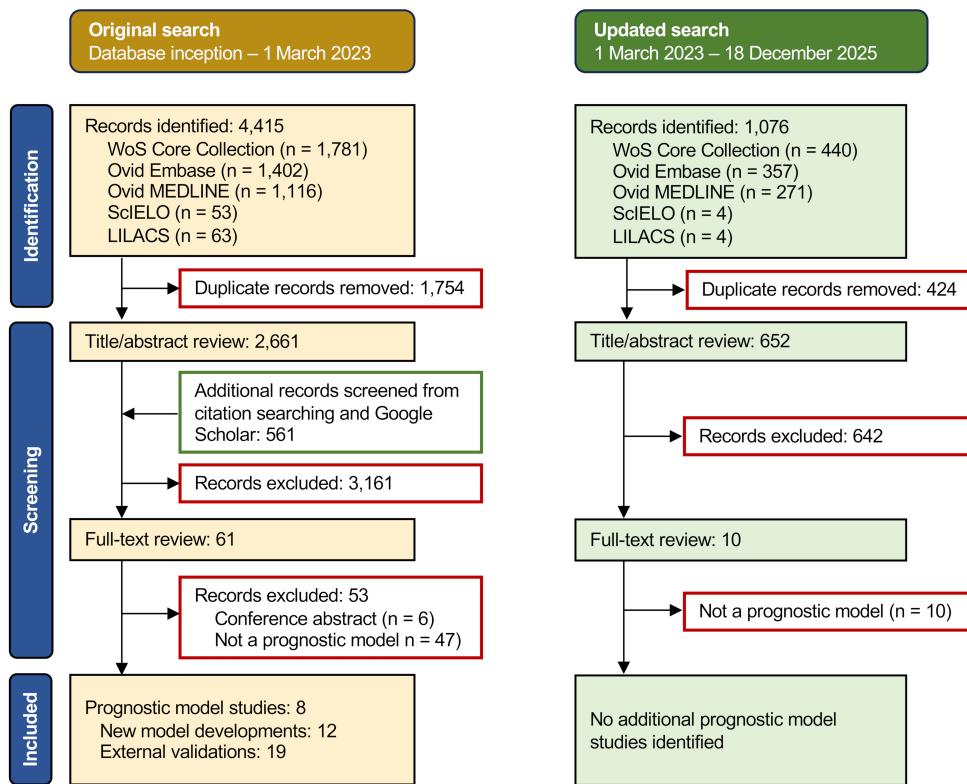


Figure 3.1: PRISMA-like flow diagram depicting the record screening process, performed initially on 1 March 2023 with subsequent updating on 18 December 2025.

3.3 Results

Study and Model Selection

After deduplication, 3,313 records were identified from the combined original and updated literature searches (Figure 3.1). Title and abstract screening yielded 71 records for full-text review, of which eight prognostic model studies were identified[163, 164, 167, 168, 183–186]. In total, 12 prognostic models were described, of which 10 underwent one or more evaluations in patient data from either different settings and/or time periods (19 external validations presented in four studies)[163, 164, 185, 186].

Model Developments

Key characteristics of the identified models are summarised in Table 3.2 (with further predictor, participant, and outcome information presented in the [Supplementary Material](#)). All models use multivariable logistic regression and predict mortality as a binary outcome, reported either as in-hospital mortality (10 models)[163, 164, 168, 184–186], or registry-reported mortality (two models)[167, 183].

Table 3.1: Glossary of key terms.

Term	Description
Prediction model	An equation or set of rules for estimating an individual's probability of an outcome based on two or more predictors. Traditionally developed using multivariable regression, although machine learning methods are increasingly used.
Outcome	The event being predicted. Also termed the response or dependent variable. Models are described as <i>prognostic</i> when outcomes occur after the time of model use, and <i>diagnostic</i> when outcomes are present at the time of model use.
Predictors	Patient or group characteristics used to estimate an outcome, also termed covariates, inputs, determinants, or independent variables. Predictors may be <i>candidate predictors</i> (considered for inclusion during model development) or <i>final predictors</i> (retained in the final model).
Overall performance	An overall summary measure of how well a model fits the data. Commonly presented measures include explained variation (R^2) and the Brier score.
Discrimination	The ability of a model to distinguish between individuals with and without the outcome. For binary outcomes this is often quantified using the concordance (c-)statistic, also termed the AUC, defined as the probability that the model assigns a higher predicted risk to an individual with the outcome than to one without. Values range from 0.5 (no better than chance) to 1.0 (perfect discrimination).
Calibration	The agreement between predicted risks and observed outcomes. For binary outcomes, calibration is best assessed using a calibration plot comparing predicted risks with observed outcome frequencies across the range of predictions.
Apparent performance	Model performance evaluated using the same dataset in which the model was developed. Performance can be optimistically biased due to overfitting, particularly in small samples or when data-driven predictor selection is used.
Internal validation	Model performance evaluated in the population represented by the development dataset, ideally using resampling techniques (e.g. cross-validation or bootstrapping) to account for overfitting. Split-sample approaches are generally considered inefficient.
External validation	Model performance evaluated in new data that were not used for model development and providing an assessment of model generalisability to new populations or settings.

Table 3.2: Key characteristics on the 12 prognostic model developments, ordered by outcome and year published. Each row corresponds to a different model developed by the referenced study. -: not reported; +ve: positive; Cand: candidate; clin: clinical; EPP: events per predictor parameter; HIV: human immunodeficiency virus; lab: laboratory; y: years

Study	Model described	Data source	Location Period	% Male	% HIV +ve	Events ¹ Sample size (%)	Predictors ² Final, Cand.	EPP ²
Outcome: Registry-reported mortality								
de Araújo 2012[183]	-	Registry	Brazil (Belo Horizonte) 2007–09	-	-	49 376 (13.0)	4, 48	1.0
Coura-Vital 2014[167]	-	Registry	Brazil (Nationwide) 2007–11	61.7	7.0	770 12,333 (6.2)	12 , (29)	(26.6)
Outcome: In-hospital mortality								
Werneck 2003[168]	-	Case-control	Brazil (Teresina) - (< 2003)	68.9	-	12 90 (13.3)	4, (15)	(0.8)
Sampaio 2010[184]	-	Retrospective	Brazil (Recife) 1996–2006	50.4	-	57 546 (10.4)	6, (15)	(3.8)
Costa 2016[164] ³	< 2 y, clin only	Prospective	Brazil (Teresina) 2005–08	-	-	- 314 (-)	6, (25)	(0.9)
	< 2 y, clin + lab	Prospective	Brazil (Teresina) 2005–08	-	-	- 291 (-)	6, (31)	(0.7)
	≥ 2 y, clin only	Prospective	Brazil (Teresina) 2005–08	-	-	- 569 (-)	9, (27)	(1.6)
	≥ 2 y, clin + lab	Prospective	Brazil (Teresina) 2005–08	-	-	- 538 (-)	9, (33)	(1.2)
Abongomera 2017[163]	-	Retrospective	Ethiopia (Abdurafi) 2008–13	95.9	19.3	99 1,686 (5.9)	8, 16	6.2
Kämink 2017[185]	< 19 y	Retrospective	South Sudan (Lankien) 2013–15	54.2	excl.	116 4,931 (2.4)	8, 20	5.8
	≥ 19 y	Retrospective	South Sudan (Lankien) 2013–15	56.2	excl.	70 1,702 (4.1)	8, 21	3.3
Foinquinos 2021[186]	Sampaio updating	Retrospective	Brazil (Recife) 2008–18	48.7	-	10 156 (6.4) ⁴	1, 1	10.0

¹ Including patients with missing predictor information and excluding patients with missing outcomes.

² Figures presented in brackets are estimated from incomplete reporting.

³ Total events in combined development dataset reported as 7.5% (66/883) and not disaggregated by age group. EPP is calculated based on extrapolating the 7.5% event rate to each model development dataset.

⁴ Sample size excludes both participants with missing predictors and missing/excluded outcomes.

Two studies (three models) were developed in East African MSF treatment centres (one model from Ethiopia that included patients with HIV/VL co-infection[12], and two models from South Sudan, for patients ≥ 19 and < 19 years, and excluding HIV/VL co-infection[185]). The remaining six studies (nine models) were performed in Brazil. Two Brazilian studies (two models) were developed using registry data; either at a national level[167] or for residents of Belo Horizonte, state of Minas Gerais[183]. The remaining Brazilian studies were developed in hospital settings, including two studies (five models) developed from patients admitted to a hospital in Teresina[164, 168], and two studies (two models) that were developed for children < 15 years and admitted to a hospital in Recife, state of Pernambuco[184, 186]. No models were developed in South Asia or the Mediterranean region.

Most studies employed a retrospective cohort design, using hospital records (four studies, five models)[163, 184–186] or registry data (two studies, two models)[167, 183]. One study used a prospective cohort design (four models)[164] and one study (one model) used a case-control design[168]. The median number of patients used for model development was 542 (range 90–12,333).

Participant age formed the inclusion criteria of eight models[164, 184–186], with five models limiting inclusion to adolescents and younger[164, 184–186]. Where reported, the median proportion of male participants was 56.2% (range 48.7–95.9%, seven models). No model excluded participants based on sex. Of the models developed in adults (≥ 15 years), patients living with HIV were either excluded (two models)[185], not reported (two models)[168, 183], and where reported, ranged from 7.0–19.3% of the model development datasets (four models)[163, 164, 167].

External Validations

Both East African studies performed external validations of their model developments (two studies, three models)[163, 185]. The model developed in Ethiopia (Abdurafi health centre, Abdurafi, Amhara region) was validated using data from a nearby treatment centre (Leishmaniasis Research and Treatment Centre, Gondar, Gondar, Amhara region)[163], and the two models developed in South Sudan were validated using retrospectively collected data from the same treatment centre (Lankien hospital, Jonglei state) and a treatment centre from a neighbouring state in South Sudan (Malakal hospital, Upper Nile state) (three external validations per model)[185].

Two Brazilian studies reported the external validations of eight models[164, 186]. One study, conducted in a prospective hospital cohort (Teresina, state of Piauí), both developed and evaluated four models. All four models were evaluated

in patients attending the same hospital over the following five years[164]. The same study used their prospective cohort to evaluate three further models: one developed from historical cohort from the same hospital[168], one developed from a retrospective hospital cohort (Recife, state of Pernambuco)[184], and one using national registry data[167]. The second Brazilian study[186] used a retrospective hospital cohort (Recife, state of Pernambuco) to both evaluate and update a model previously developed in the same hospital[184].

Further details on validation datasets are presented in the [Supplementary Material](#).

Model Performance

Model discrimination measures were reported as c-statistics for all risk scores, and presented in Table 3.3. Further performance measures, where reported, are presented in the [Supplementary Material](#). For external validations, the median c-statistic was 0.78 (range 0.62–0.92, 10 models, 19 external validations). When evaluated in the same patients used for model development (apparent performance), the median c-statistic was 0.86 (range 0.56–0.93, 12 models). No studies presented overall measures of performance or calibration plots. One model's calibration plot could be reproduced from an internal (split-sample) validation of the risk score[167].

Table 3.3: Summary of model discrimination and risk of bias assessments, ordered by outcome and year published. Each row corresponds to a individual model that was either developed or externally validated in the referenced study. Unless otherwise stated, the reported c-statistics represent performance of the model risk score and not the full model equation. -: not reported; dev: development; eval: evaluation; n/a, not applicable (external validations only); perf: performance; val: validation; **Risk of bias (RoB):** A: analysis, O: outcome; OA: overall assessment; P: participants; Pr: predictors; +: high RoB; -: low RoB; ?: unclear RoB

Study	Model described	Model discrimination (c-statistic) ¹			Risk of bias ²				
		Apparent perf.	Internal val.	External val.	Eval type	P	Pr	O	OA
Outcome: Registry-reported mortality									
de Araújo 2012[183]	-	0.756	-	-	dev	+	+	?	++
Coura-Vital 2014[167]	-	0.80 (0.78–0.82)	0.78 (0.75–0.82) ³	-	dev	+	+	?	++
Outcome: In-hospital mortality									
Werneck 2003[168]	-	0.882	-	-	dev	+	+	-	++
Sampaio 2010[184]	-	0.895	-	-	dev	+	+	-	++
Costa 2016[164]	< 2 y, clin only	0.90 (0.84–0.97)	-	0.83 (0.64–1) 0.86 (0.74–0.98)	dev val (×2)	-	-	-	++
	< 2 y, clin + lab	0.93 (0.88–0.98)	-	0.80 (0.57–1) 0.92 (0.84–1)	dev val (×2)	-	-	-	++
	≥ 2 y, clin only	0.89 (0.84–0.93)	-	0.75 (0.68–0.83) 0.88 (0.83–0.93)	dev val (×2)	-	-	-	++
	≥ 2 y, clin + lab	0.92 (0.88–0.96)	-	0.79 (0.62–0.96) 0.71 (0.34–1)	dev val (×2)	-	-	-	++
	Werneck 2003	n/a	n/a	0.75	val (×1)	?	-	-	++
Abongomera 2017[163]	Sampaio 2010	n/a	n/a	0.87	val (×1)	?	-	-	++
	Coura-Vital 2014	n/a	n/a	0.77	val (×1)	?	-	-	++
	-	0.83 (0.79–0.87)	0.82 (0.77–0.88) ⁴	0.78 (0.72–0.83)	dev val (×1)	+	-	-	++
Kämink 2017[185]	< 19 y	0.83 (0.78–0.87)	-	0.72; 0.83; 0.77	dev val (×3)	+	+	-	++
	≥ 19 y	0.74 (0.68–0.81)	-	0.72; 0.80; 0.71	dev val (×3)	+	+	-	++
	Foinquinos 2021[186]	Sampaio updating	0.556 0.762 (0.662–0.901) ⁵	-	dev	+	+	-	++
Foinquinos 2021[186]	Sampaio 2010	n/a	n/a	0.618 ⁵	val (×1)	+	+	-	++

¹ Significant figures and 95% confidence intervals are reproduced as reported in the publication.

² For studies presenting multiple external validations of the same model, risk of bias assessments were the same and therefore grouped together.

³ Split-sample (random split with 2:1 development:validation).

⁴ Cross-validation (5-fold).

⁵ C-statistic from evaluation of full model equation.

Predictors

A visual comparison of the candidate and final predictors is presented in Figure 3.2. Where authors provide further predictors definitions, these are reported in the [Supplementary Material](#). The four most common candidate predictors were jaundice (12 models), age (11 models), sex (10 models) and bleeding (10 models). Initial VL treatment was included as a candidate predictor in two models, although not retained in the final models [12,35]. Predictors most frequently retained in the models were jaundice (11 models), bleeding (eight models), and age (seven models). No models included sex as a final predictor. One model did not include HIV status as a candidate predictor, despite being conducted in adults and not pre-specifying the exclusion of patients with HIV [20]. Apart from HIV testing, four models did not consider laboratory tests as a predictor[[164](#), [167](#), [183](#)]. The remaining eight models included laboratory tests both as candidate and final predictors[[163](#), [164](#), [168](#), [184–186](#)].

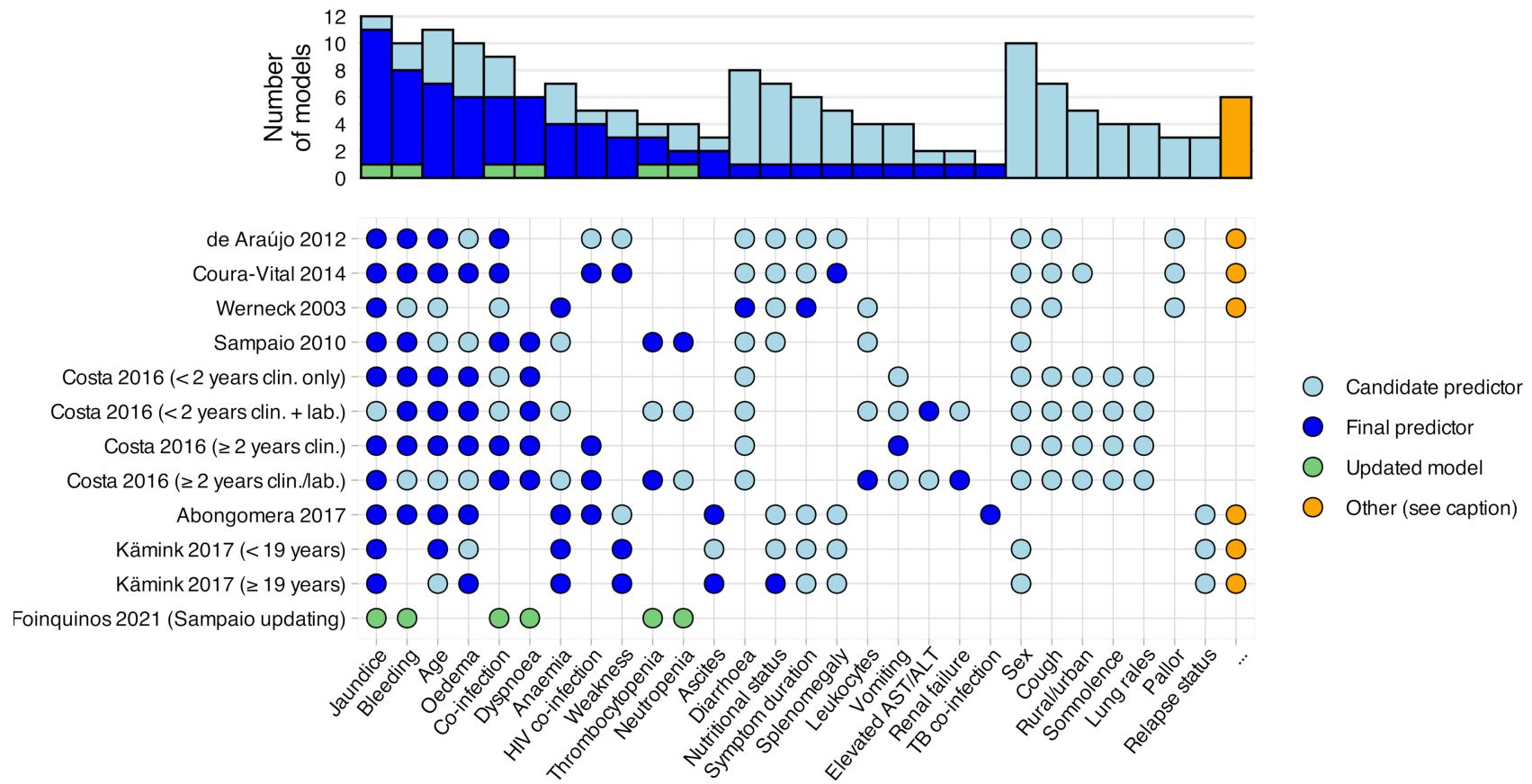


Figure 3.2: Candidate (considered) and final (retained) predictors included in prognostic models of mortality in visceral leishmaniasis. Bars indicate the number of models incorporating each predictor. Models are labelled by first author, publication year, and model name (in brackets where multiple models were reported). Conceptually similar predictors were grouped and renamed; definitions and groupings as reported by the study authors are provided in Supplemental Table 4. For models by de Araújo et al. and Coura-Vital et al., cough and/or diarrhoea were treated as a single predictor in accordance with national registry reporting. For models by Kämink et al., oedema and ascites were combined as a single predictor, whereas Abongomera et al. treated these as separate variables. Predictors shown for the updated model correspond to the predetermined final predictors of the model being updated (Sampaio et al.). Abbreviations: ALT, alanine transaminase; AST, aspartate transaminase; clin, clinical; HIV, human immunodeficiency virus; lab, laboratory; TB, tuberculosis. ‘..’ predictors assessed in ≤ 2 models and not included in the final model: [Werneck 2003]: ‘abdominal distension’, episodes of blood transfusion; [de Araújo 2012]: fever, hepatomegaly, “other clinical manifestations”, initial VL regimen, VL drugs following initial regimen, antimony treatment duration; [Coura-Vital 2014]: fever, hepatomegaly, ‘other clinical manifestations’, race, education; [Abongomera 2017]: initial VL regimen; [Kämink 2017, both models]: lymphadenopathy.

Model Presentation

All 12 models were presented as simplified risk scores. Score ranges, suggested risk groupings, and the authors' opinions on how different risk groups should inform clinical decision-making are presented in the [Supplementary Material](#). Outcome (mortality) probabilities corresponding to the risk scores were presented either in tabular format (four models)[[163](#), [167](#), [185](#)], graphically and through a web application (four models)[[164](#)], or not presented (four models)[[168](#), [183](#), [184](#), [186](#)]. The full model equation, including model intercept, was reproducible for three models in total — presented either in the original study describing the model development (one model)[[167](#)], or in the updating study (two models, corresponding to the original model and updated model)[[186](#)].

Risk of Bias Assessment

Risk of bias assessments for all model developments and external validations are presented in Table [3.3](#), alongside measures of discrimination and details of model presentation and reproducibility. Further details on variable selection and handling of missing data are presented in the [Supplementary Material](#). Responses to the risk of bias signalling questions are also provided in the [Supplementary Material](#).

All 12 model developments were judged at an overall high risk of bias.

The analysis domain was assessed at high risk of bias across all model developments. One model obtained a sufficient sample size (event to predictor parameter ratio > 10), and adequately reported model performance, including calibration[[167](#)]. All models (excluding model updating) were developed using a univariable selection stage and did not adjust model predictions to account for optimism due to overfitting. In four model developments, the presented risk scores were not reproducible from the regression model coefficients(full calculations are presented in the [Supplementary Material](#))[[164](#)].

Five models (two studies) were assessed as having a low risk of bias in the predictors domain[[163](#), [164](#)]. Both studies provided evidence that the model predictors were defined consistently for all patients and were assessed without knowledge of outcome data. The remaining seven models (six studies) were considered at high risk of bias, with one model including predictors that were likely measured after the time of intended model implementation[[183](#)].

The outcome domain was assessed at low risk of bias for all model developments, except for two models where bias risk was unclear[[167](#), [183](#)].

The participants domain was assessed at low risk of bias for four models (one study)[164], with the remaining models considered high risk due to using retrospectively collected data.

All 19 external validations were also judged at high risk of bias, although assessment across the domains was limited by a lack of reporting. Briefly, sources of bias were similar to those identified for model developments, including small sample sizes, the use of retrospectively collected data, and the absence of calibration measure reporting. Please refer to the [Supplementary Material](#) for further details.

3.4 Discussion

Across a range of diseases, the number of prediction model studies has surged over the last two decades, driven by an increasing focus on personalised medicine, the need to provide evidence for guideline development, and a growing number of tools available for model development[103, 160, 187]. VL is no exception, with a total of 12 prognostic models identified, of which nine were published since 2013. All models were developed in Brazil or East Africa and predict either in-hospital or registry-reported mortality.

When using a prognostic model to predict mortality, for example, in a hospital ward or outpatient clinic, the clinician should be confident that after inputting model predictors (for example, age, haemoglobin, clinical signs and symptoms), they receive a trusted estimate of the probability (risk) of the outcome occurring. Empowered with this information, the patient can then be counselled, and important treatment decisions agreed on. Having confidence in the model output is fundamental, since inaccurate risk predictions can lead to suboptimal decision-making, inequitable care, and at times, patient harm.

3.4.1 Prognostic Model Assessment

Three important considerations should be taken into account when assessing whether a prognostic model's estimated risks are reliable for a target patient. These include (1) applicability: whether the model has been developed or evaluated in a setting and population similar to that of the target patient; (2) performance, including both discrimination and calibration, in the populations and settings chosen for evaluation; and (3) risk of bias: whether the studies performing model development or evaluation are subject to systematic errors that may distort risk estimates and performance measures.

Applicability

First, we consider model applicability. Has the model been developed, and/or evaluated in patients similar to the patient I'm interested in? Here, the onus is on the model user to compare their target patient and setting to those in which the model was developed and/or evaluated (Table 3.2, [Supplementary Material](#)). In VL, important mismatches can result from differences in (i) HIV status, (ii) patient age, (iii) geographical setting (both locally within a country, and between endemic regions), (iv) treatments used (where reported), (v) temporal differences, and (vi) treatment setting (e.g. inpatient vs outpatient). For example, estimating the mortality risk of a patient with HIV co-infection using a model developed from HIV-negative patients is likely to underestimate the true risk. Similarly, using a model developed in Brazil to estimate mortality risk in patients from India, may well overestimate the true risk, given the overall lower mortality rate in South Asia[3]. One may also question the contemporary risk estimates of models developed using data from 10-20 years ago. For example, is it fair to assume that models developed using hospital or registry data in Brazil in the 2000s and early 2010s are still accurate, given significant changes in treatment, and an evolving disease epidemiology[188]? Questions regarding model applicability, such as these, often have no clear consensus answer, although are important to consider.

Performance

Second, even if a model were applicable to my patient population, how can I be sure it performs well? As we summarise here (Table 3.3 and [Supplementary Material](#)), model discrimination, presented as the c-statistic, is universally reported across all identified models, both when evaluated in the same patients used for model development (apparent performance) and when evaluated in new patients (external validation). Furthermore, given the reported c-statistics are frequently over 70–80%, even when evaluated in new data, we are reassured that in the appropriate population, most models do a fairly good job at assigning higher risk to those who progress to death compared to those who do not. Crucially, while a highly discriminative model can reliably rank patients by risk, discrimination alone provides no information on the accuracy of the predicted probabilities. A model that estimates a mortality risk of 20%, where the true risk is 2%, may still have 100% discriminative performance, despite dramatically overestimating the true risk.

Instead, we should be reviewing a model's calibration, ideally presented as a plot of observed vs. estimated risks, to assess this fundamental, yet frequently overlooked

aspect of model performance[189]. However, with estimated and observed risks only reproducible for one model risk score[167], we find that the VL prognostic model landscape bucks the broader trend: calibration is neglected[159, 169, 190]. In contrast, measures of discrimination and classification (sensitivity, specificity), are preferentially reported, despite their limited clinical utility[173].

Risk of Bias Assessment

Finally, we consider risk of bias, alongside the closely related issue of model reporting, both of which directly influence the interpretability, reproducibility, and clinical use of prognostic models.

All model development and external validation studies were judged to be at high risk of bias according to PROBAST. Importantly, this classification does not imply that the identified models are inherently flawed or clinically uninformative. Rather, PROBAST highlights aspects of study design, analysis, or reporting that can challenge interpretation of estimated risks and performance measures[173]. Below, we briefly describe several recurrent issues and direct readers to accessible guidance on best practice in prognostic modelling.

A frequent concern was model overfitting, which occurs when models capture random variation specific to the development dataset rather than true underlying risk patterns. This is more likely when many predictors are considered relative to the number of outcome events, and can lead to exaggerated performance estimates and inflated risk predictions when models are applied to new patients[191, 192]. While rules of thumb often suggest a minimum of 10–20 events per predictor parameter, most identified models fell well below this threshold. Formal sample size calculations are now available for prediction model development, which can also be used to assess the number of predictors that can be reliably supported for a given sample size[46].

Internal validation provides a means of adjusting performance estimates and risk estimates for the effects of overfitting. However, only two models applied internal validation methods, either using cross-validation[163] or a split-sample approach[167]. Although historically common, data splitting is increasingly discouraged, particularly in smaller datasets, as it reduces the effective sample size available for model development. Resampling methods, such as bootstrapping, allow full use of the available data while providing more reliable estimates of model performance[158].

Another recurring source of bias was predictor selection based on univariable analyses. This approach can result in unstable models and misleading predictor inclusion, as variables are selected based on isolated statistical associations rather

than their joint contribution to risk prediction. Alternative strategies include pre-specifying predictors based on clinical relevance, or applying penalisation or dimension-reduction techniques to limit model complexity[173, 193].

We also identified reporting-related concerns that directly affected reproducibility of derived risk scores, with one study (four models) presenting scores that did not correspond to reported regression coefficients. While these discrepancies may reflect reporting errors, they undermine confidence in score implementation and external use.

Other concerns related to model reporting and reproducibility were commonplace, limiting both model appraisal and application to the individual patient. Notably, several studies did not report absolute risk estimates corresponding to the presented risk scores[168, 183, 184, 186]. When models are presented primarily as binary classifiers (e.g. high vs low risk), much of their potential clinical utility is lost, as clinicians are unable to interpret or communicate individualised risk estimates.

Similarly, incomplete presentation of the full model equation—including the intercept term—restricted reproducibility and external evaluation. Without this information, users are limited to applying simplified risk scores, which may differ in performance from the underlying regression model[173]. As only two studies reported the full model equation[167, 186], independent evaluation of the remaining models would require direct contact with the original authors.

Additional sources of bias—such as categorisation of continuous predictors, uncertain or suboptimal handling of missing data, reliance on retrospective data sources, and model evaluation in datasets with few outcomes—were also common. Importantly, the issues identified in this review are not unique to VL, but mirror challenges repeatedly highlighted in systematic reviews of prognostic models across a wide range of disease areas[103, 159, 173, 194, 195]. For readers interested in addressing these issues, we signpost accessible and authoritative guidance on best practices in model development[193], evaluation[196, 197], sample size calculation[198, 199], model reporting[103], model presentation, (including risk score derivation)[182], and updated guidance on risk of bias assessment[200].

3.4.2 Predictors of Mortality

Where outcome timing was reported, death frequently occurred within days of hospitalisation, indicating a short prediction horizon. Two of the included prediction model studies reported time to death: one-third of patients died within 48 hours of admission in South Sudan[185], and the average time to death was just over five days in a study reporting from Teresina (Piauí, Brazil). In this context,

predictors retained across models predominantly reflect advanced disease and imminent physiological decompensation.

Severe VL is increasingly understood as a progressive inflammatory syndrome — described as ‘leishmanial sepsis’ and characterized by cytokine storm, disseminated intravascular coagulation, secondary bacterial sepsis, and evolving multiple organ dysfunction[51]. It is therefore unsurprising that frequently identified predictors of mortality—including jaundice, bleeding, dyspnoea, oedema, and bacterial co-infection (cf. Figure 3.2) represent clinical markers of established organ failure and broadly align with factors identified in systematic reviews of prognostic factors for VL mortality in East Africa[201] and Brazil[202].

Conversely, factors plausibly important earlier in the disease course, such as symptom duration, nutritional status, or relapse history, were infrequently selected. While these variables may influence susceptibility or delayed care-seeking, they appear to have limited discriminatory value once patients reach hospital with severe disease.

The short prediction horizon also explains the relatively high discrimination reported across studies, with c-statistics frequently approaching or exceeding 0.85. Under these conditions, discrimination is driven by late-stage clinical features rather than early prognostic signals. As a result, existing models primarily identify patients at high risk of near-term mortality, rather than supporting earlier risk stratification or anticipation of clinical deterioration.

3.4.3 Implications for Future Research

Several important evidence gaps were identified in this review. Most notably, no prediction model studies were identified outside East Africa and Brazil. This is striking given that South Asia has, until recently, accounted for the lion’s share of the global VL burden. Mortality rates in South Asia are, however, relatively low compared with Brazil and East Africa, meaning that relatively large sample sizes would be required to develop mortality models without substantial risk of overfitting. In this setting, routinely collected programme data and national surveillance systems may offer a pragmatic opportunity to support adequately powered model development, provided data quality and outcome ascertainment are sufficient.

A further major evidence gap is the absence of models predicting relapse or post-kala-azar dermal leishmaniasis. These outcomes are of particular importance in elimination settings, since both represent persistent infection reservoirs that sustain transmission[85, 203]. The lack of current prognostic tools in this area limits our ability to target follow-up, secondary prophylaxis, or intensified surveillance.

Recently the WHO released a target product profile for a test of cure following treatment, such that patients at high risk of subsequent relapse can be identified early[88]. A prognostic model for relapse could serve as a surrogate for such an in vitro test. Leveraging IPD meta-analysis allows harmonisation of data across heterogeneous studies, increases statistical power, and maximises the value of existing datasets—particularly in disease areas where high-quality data are scarce.

In Brazil, existing models — including those informing current national guidelines[60] — are based on data collected 10–20 years ago. Since that time, the epidemiology of VL has evolved, and treatment practices have shifted substantially, for example, with higher rates of HIV/VL co-infection, an increasing age at diagnosis, and better access to liposomal amphotericin B[188]. These changes raise questions about the continued validity of older models and highlight the need for model evaluation with recent patient data and, where necessary, model updating. In addition, future studies should carefully consider whether it is appropriate to combine patients with and without HIV within a single model, given their distinct clinical trajectories, immunological profiles, and risk factors for adverse outcomes[106].

Finally, emerging applications of artificial intelligence and machine learning warrant cautious consideration in the VL prediction modelling landscape. Recent studies have demonstrated the feasibility of applying machine learning methods to VL datasets[204, 205]. However, as emphasised in PROBAST+AI and TRIP-POD+AI guidance, such approaches are not immune to bias, overfitting, or poor transparency[103, 200]. Without rigorous reporting, external validation, and explicit consideration of clinical use-cases, machine learning models risk offering limited real-world utility. Future research should therefore prioritise methodological robustness, interpretability, and clinical relevance over algorithmic complexity.

3.4.4 Limitations

The principal limitations of this review relate to its scope. We excluded unpublished manuscripts, including conference abstracts and educational theses, due to concerns about variable methodological quality and access; however, this may have led to the omission of emerging or ongoing work.

In line with our prespecified inclusion criteria, we also excluded studies reporting diagnostic models[206, 207], models that were not developed using a multivariable approach[143, 151, 208], prognostic factor studies[13, 114, 209, 210], and registered but unpublished prognostic model studies[211]. While these exclusions were necessary to maintain a focussed research question, they limit our ability to provide a comprehensive overview of the full VL prediction research landscape.

3.4.5 Conclusion

We present the first systematic review that identifies, summarises, and critically appraises prognostic models for all VL clinical outcomes. Using established methodological guidance, we provide a comprehensive, objective, and transparent synthesis of model characteristics, performance, applicability, and risk of bias. Our findings highlight substantial gaps in the current evidence base.

All identified models predict mortality, were developed exclusively in Brazil or East Africa, and are almost all based on data collected over a decade ago. No models were developed in South Asia or the Mediterranean region, nor addressed relapse or PKDL — outcomes that are central to elimination efforts. These represent important evidence gaps where future research efforts should be prioritised.

Clinicians, researchers, and policymakers can refer to this review to assess the strengths and limitations of existing VL prognostic models in this highly neglected and often fatal disease.

We direct interested readers to expert guidance to support transparent reporting and reduce common sources of bias in the development and evaluation of prediction models.

4

Methodology

Contents

4.1	Introduction	51
4.2	Data Harmonisation	53
4.2.1	Data Acquisition	54
4.2.2	Data Curation	54
4.2.3	Population at Risk	56
4.2.4	Outcome	59
4.3	Model Development	60
4.3.1	Sample Size	60
4.3.2	Candidate Predictors	62
4.3.3	Descriptive Analysis	66
4.3.4	Missing Data	67
4.3.5	Model Specification	70
4.3.6	Variable Selection	71
4.3.7	Model Performance	72
4.4	Internal Validation	74
4.5	Summary	75

4.1 Introduction

As outlined in Chapter 2, VL relapse is consequential not only for individual patients but also poses a threat to sustained elimination efforts. Accordingly, the development of a non-invasive tool to predict relapse — functioning as a ‘test-of-cure’ following initial treatment — has been identified by the WHO as a research priority[86, 88, 203]. A prognostic model represents a potential solution: by

quantifying the relationship between patient characteristics and subsequent relapse events, relapse risk in future patients can be estimated and clinical decision-making informed. However, as demonstrated in Chapter 3, no prognostic models for VL relapse have been published to date.

To address this evidence gap, four prognostic models are developed using IPD from the IDDO VL data platform: two for patients from the ISC and two for patients from East Africa. Within each region, one model includes parasite grade at baseline (pre-treatment) assessment, and one model excludes parasite grade, reflecting differences in data availability and clinical practice. All models use routinely collected information available at the time of initial cure assessment to predict six-month relapse among VL patients without HIV co-infection.

The development and evaluation of clinical prediction models is supported by an extensive and growing methodological literature. Over the past decade, reporting guidelines for prediction model studies have been established[103, 158, 176, 212], alongside an increasing number of reviews and recommendations that define best practice[193, 196, 197, 199, 213]. In particular, the application of meta-analysis techniques to IPD from multiple studies presents exciting new opportunities for prediction model development and evaluation[212, 214]. Notable opportunities include increased sample sizes leading to greater statistical power, and the ability to explore heterogeneity in predictor effects and model performance across different settings. Additionally, IPD can be used to standardise inclusion criteria and outcome/predictor definitions across included studies. However, as we lay out in this chapter, the use of IPD in prediction model research also introduces challenges; specifically (i) the need for statistical models that account for clustering of participants within studies and (ii) the presence of missing data, which can be sporadically missing within studies, or entirely missing from one or more studies[212].

The aim of this chapter is to describe and justify the methodology used for model development and evaluation — from data acquisition to final model presentation. Guidelines and methodological texts are cited accordingly, and checklists provided for current reporting guidelines on prediction model studies (TRIPOD-AI and TRIPOD-Cluster)[103, 212]. Additional material is presented in Appendix B and in the Supplementary Material.¹ A protocol is available on the Open Science Framework.²

This chapter's structure closely mirrors the methodological workflow as outlined in Figure 4.1, with sections on data harmonisation, model development, and internal validation. In keeping with best practice, and similar to the approach

¹Available at <https://github.com/jpwil/dphil>.

²Created Nov 8, 2024, available at <https://osf.io/z4bdn>.

adopted in Chapter 3, the research question is presented in Box 4.1 using the PICOTS (population, index model, comparator model, outcome, timing, and setting) framework[174, 214]. Further elaboration of the eligibility criteria, and standardised definitions of predictors and the outcome are considered in the following section.

All analyses were performed using R version 4.4.1[215], with R packages cited in the relevant sections below. R scripts used for model development and evaluation are provided in the [Supplementary Material](#).

Box 4.1: Definition of the research question: a PICOTS approach

Population HIV-negative patients that are prospectively recruited into a clinical trial with a diagnosis of visceral leishmaniasis, confirmed either serologically or parasitologically. No restrictions are placed on age, sex or treatment regimen.

Index models For each setting, two prognostic models are developed; one including baseline parasite grade from a tissue aspirate, and one without. The models predict the *future* occurrence of relapse, using patient information collected at treatment baseline. The intended time of model use is following a successful assessment of treatment response.

Comparator model As established in Chapter 3, no published relapse models are available for comparison or updating.

Outcome Relapse is defined as the recurrence of signs and symptoms of VL requiring rescue treatment, and following demonstration of an initial treatment response.

Timing Relapse occurring within 6-months of test-of-cure (typically occurring at the time of treatment completion, or within 30 days of starting treatment).

Setting Participants from either the Indian subcontinent or East Africa.

4.2 Data Harmonisation

Here, data harmonisation refers to the process of data acquisition, curation, and any subsequent data manipulation required to produce a single analysis dataset ready for model development.

In the interest of full disclosure, data acquisition was completed by IDDO colleagues prior to the commencement of this DPhil project. The first stage of data curation — conversion of the contributed datasets to the Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) standard — was performed by the IDDO data engineering team with support from the IDDO science team (myself included). Subsequent methodological steps were led by myself.

4.2.1 Data Acquisition

A systematic review of the scientific literature was first performed in 2016, with the aim of comprehensively cataloguing all existing VL clinical trials (PROSPERO: CRD42021284622)[216]. 145 trials were initially identified (1980–2016, n = 26,986 patients), with further trials added during periodic updates according to an open protocol[217]. Between 2018–2022, corresponding authors of the identified VL clinical trials were invited to share their IPD with the IDDO VL data platform, in line with the General Data Protection Regulation (GDPR)-compliant IDDO data sharing governance[218, 219].

4.2.2 Data Curation

Conversion of the contributed datasets to an analysis-ready dataset of all eligible IPD occurred in two key stages.

Stage 1: CDISC SDTM Curation

To facilitate reusability and interoperability, contributed datasets were standardised to a common storage format: the CDISC-compliant SDTM standard[220], adapted by IDDO for VL[221]. During this process, contributed datasets underwent *pseudonymisation*,³ prior to being available for data sharing requests. Briefly, SDTM format datasets comprise a number of standardised domains (tables) containing related information (e.g. patient demographics, laboratory results, treatment administration, clinical signs and symptoms). Each domain contains a set of standard variables (table columns, e.g. STUDYID, USUBJID, VISITDY) alongside VL-specific variables defined by IDDO (e.g. parasite grade, spleen size). Further details of the curation process are available in the [IDDO SDTM Implementation Guide](#).⁴

³ ‘...processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information’ <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/pseudonymisation/> (accessed 15 Dec 2025).

⁴ Available at <https://www.iddo.org/tools-and-resources/data-tools>, free registration required.

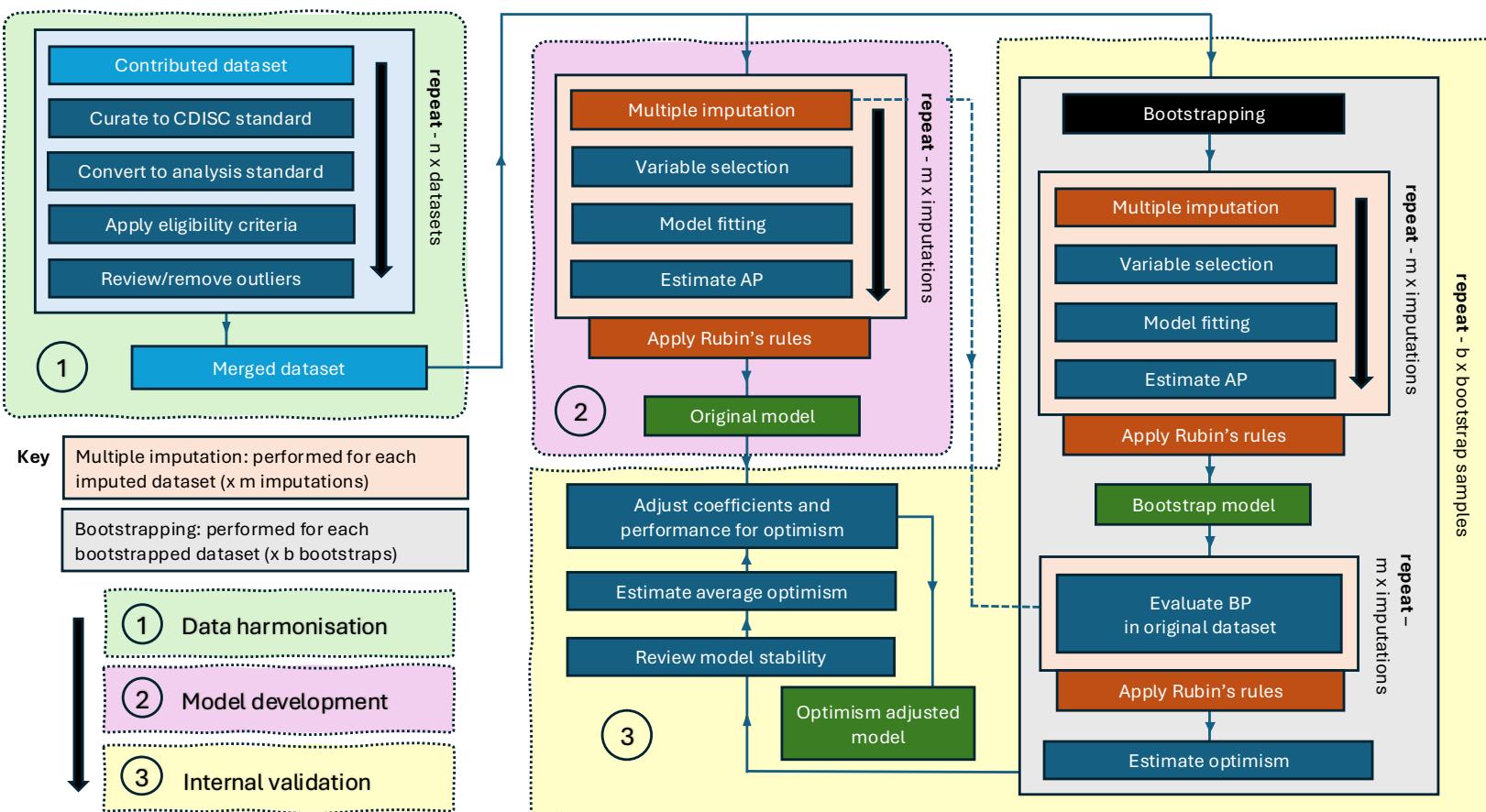


Figure 4.1: Schema of methodological workflow. ① Data harmonisation is performed for each contributed dataset including: initial curation (by the IDDO data engineering team) to the CDISC SDTM format, application of inclusion and exclusion criteria, and removal of outliers. Curated and cleaned datasets are converted into an analysis (wide) format and merged prior to ② model development. Multiple imputation is used to create ($m = 30$) imputed (complete) datasets. Variable selection, model fitting, and apparent performance evaluation is performed on all imputed datasets. Estimates are pooled using Rubin's Rules. Bootstrapping is used to perform ③ internal validation, allowing (i) review of model stability and (ii) optimism adjustment of performance measures and original model coefficients. All model development steps, including multiple imputation, are repeated for each of the ($b = 500$) bootstrap datasets. The resulting bootstrap models ($b = 500$) are evaluated both in the corresponding bootstrap (imputed) datasets and the original (imputed) datasets, and pooled using Rubin's Rules. The mean of the differences of the pooled performance measures (in bootstrap vs. original dataset) are used to shrink the original model coefficients and apparent performance measures, resulting in the final optimism-adjusted model. AP: apparent performance; b: number of bootstraps; BP: bootstrap model performance; CDISC: Clinical Data Interchange Standards Consortium; IDDO: infectious diseases data observatory; m: number of imputations; n: number of contributed studies. SDTM: Study Data Tabulation Model.

Stage 2: Analysis-Ready Dataset Curation

Subsequently, SDTM format datasets were converted to a single analysis-ready dataset, primed for model development. This stage consisted of multiple steps, refined iteratively over several months and in close consultation with the IDDO data engineering team:

- Identification and removal of spurious data points (e.g. outliers, discussed below)
- Application of study and participant eligibility criteria (Section 4.2.3)
- Creation of a standardised outcome variable according to a pre-defined definition (Section 4.2.4)
- Conversion of the datasets from a long to wide format, consisting of one row per participant
- Merging of all datasets into a single analysis dataset

Data wrangling during the second curation stage was performed with the `tidyverse` suite of R packages[222]. Identification and removal of spurious data points was performed through subgroup tabulations and visual inspection of histograms and scatter plots. Where two incongruous data points were identified, for example, incompatible height and weight values, a third variable, such as BMI or age, would be used to identify the spurious value. Data points considered to be outliers were converted to missing values. A complete record of all data cleaning steps, including outlier identification and removal, was maintained and documented in commented R scripts.

In Section 2.3 of Chapter 2, relapse was defined broadly as ‘the reappearance of VL signs and symptoms following an initial treatment response’, and ‘typically confirmed by direct visualisation of the parasite on a tissue aspirate smear’. Despite appearing a clear definition, on closer inspection it can be appreciated that even subtle variations in eligibility criteria, study design, and the definition of efficacy endpoints can, at times unexpectedly, impact the proportion of patients experiencing relapse as a study outcome.

4.2.3 Population at Risk

Clear specification of the population at risk is fundamental to understanding the model’s real-world applicability[103, 214]. Particular attention is given to the definition of initial cure, since (i) relapse can only occur following an initial

Box 4.2: Eligibility criteria

- **Study-level inclusion criteria**
 - Studies conducted in either the ISC (India, Nepal, Bangladesh) or East Africa (Ethiopia, Sudan, South Sudan, Kenya, Uganda)
 - Prospective design, defined as participants having provided informed consent
 - Participants recruited with a diagnosis of VL as defined by a combination of clinical symptoms and either parasitological or serological confirmation
 - Studies that report, as a minimum, the treatment regimen including at least the drug name(s), dose and duration
 - Recruited a minimum of 6 patients
 - Included a minimum of 6 months of prospective follow-up from treatment initiation
 - Reported VL relapse events during the 6-month follow-up period
- **Participant-level exclusion criteria**
 - Participants with HIV co-infection or from a setting with high HIV co-infection prevalence and without a negative HIV test
 - Participants who were confirmed pregnant at the time of treatment initiation
 - Participants with symptomatic treatment failure requiring rescue treatment, identified either before or at initial cure assessment

treatment response, and (ii) heterogeneity in study-level cure definitions can be partly addressed through IPD-based standardisation.

Inclusion and exclusion criteria were applied at the study and participant levels and are presented in Box 4.2. Criteria are chosen according to (i) the eligibility criteria applied in the original systematic review from which identified study authors were invited to contribute their IPD[216], (ii) the range of studies available in the IDDO VL data platform, and (iii) the resulting impact and applicability of models developed.

Study-Level Criteria

Study-level inclusion criteria were applied to ensure that contributed studies were sufficiently comparable in terms of epidemiological context, study design, and outcome ascertainment to permit meaningful harmonisation and pooled analysis.

Studies were limited to those conducted in the ISC and East Africa, reflecting both the public health relevance of relapse prediction in regions with ongoing VL

elimination programmes, and the availability of IPD the IDDO VL data platform. On review of the IDDO inventory[223], only two studies were conducted outside these regions — one in Greece conducted in the 1990s [224], and one in Brazil in the 2010s [225]. These were excluded to preserve geographical and epidemiological coherence.

Only prospectively conducted studies were included. Prospective designs allow for systematic and active follow-up, predefined outcome definitions, and contemporaneous outcome recording, all of which are important for the reliable identification of initial cure and subsequent relapse. However, reliance on IPD from clinical trial settings limits model applicability to real-world patients — those who are managed outside trial settings, and may not meet the often-stringent eligibility criteria. These limitations are discussed further in Chapter 7.

A minimum study size of at least 6 patients was imposed to exclude very small cohorts with unstable relapse estimates. Finally, studies were required to report relapse events during follow-up, either explicitly or in a form that allowed relapse to be inferred from the available IPD.

Participant-Level Criteria

With respect to clinical presentation, treatment response, and outcomes, patients with VL/HIV co-infection constitute an important but distinct clinical population. Accordingly, patients with and without VL/HIV co-infection were *not* combined within a single prediction model, given the substantial uncertainty in extrapolating relapse associations derived from HIV-negative patients to those with VL/HIV co-infection. Since the majority of contributing studies excluded patients with VL/HIV co-infection, insufficient IPD were available to develop a separate model for this group. Patients without evidence of HIV testing were excluded if they were otherwise considered to be from a high risk area. For the East Africa models, patients without HIV testing were only included if they were recruited from Sudanese sites.

As with VL/HIV co-infection, very few contributing studies included pregnant participants, reflecting their systematic exclusion from VL trials and precluding the development of a separate prediction model.

Initial Cure

Understanding study-specific definitions of initial cure is important, as all studies require the patient to demonstrate a treatment response, measured with a test-of-cure, in order to be at subsequent risk of relapse. Consequently, patients *not* achieving initial cure — described as initial treatment failure — should be excluded.

A direct consequence of excluding these patients is that model-derived risk estimates are only applicable to patients demonstrating initial cure.

Initial cure definitions based solely on clinical improvement, as is common in routine practice, are likely to classify some patients as cured despite persistently positive tissue aspirates, were these assessed. These patients form a subgroup at increased risk of relapse and would instead be classified as initial treatment failures under more stringent, parasitology-based test-of-cure criteria, thereby being excluded from subsequent follow-up. Consequently, all else being equal, studies applying stricter definitions of initial cure will observe a lower subsequent relapse risk.

Recently, Dahal and colleagues performed a systematic review of the design, conduct, analysis, and reporting of VL therapeutic efficacy studies, published between 2000 and 2021. Of the 89 studies identified, 71 (79.8%) included parasitological assessment, with or without demonstration of clinical improvement, as part of the test-of-cure, while 13 (14.6%) required clinical improvement only. The remaining studies did not provide a definition. Timing also varied considerably, with the 68 (76.4%) of studies performing the test-of-cure between 15–30 days following treatment completion[90]. Similar patterns are observed in the contributed studies, as reported in the [Supplementary Material](#), and discussed further in subsequent results chapters. Importantly, criteria for ‘clinical improvement’ are often not specified. Further complicating interpretation, many studies describe a subgroup of ‘slow responders’, who remain in the study despite a positive tissue aspirate in the test-of-cure. These patients may undergo repeat assessment at subsequent time points (e.g. 2–4 weeks later), with or without treatment extension, and may or may not ultimately be classified as having achieved initial cure.

Such variation in the initial cure definition can challenge standardisation efforts, leading to differences in observed relapse rates stemming from differences in the population at risk. These differences, however, can often be mitigated through interrogation of the IPD. Box 4.3 provides a working definition for initial cure, which is applied during data harmonisation.

4.2.4 Outcome

Relapse itself, where described at the study-level, is also subject to substantial variation with respect to its (i) definition — including the severity of symptoms required to trigger a repeat aspirate and the tissue type chosen for aspirate, and (ii) timing — including whether patients were actively screened at set time points with clinical examination ± routine aspirates, or whether dependent on patients attending voluntarily based on recurrent symptoms and discharge advice. In line

Box 4.3: IPD-based working definitions

Initial cure Where initial cure (or initial treatment failure) is *not* directly recorded in the IPD as an efficacy outcome, or where it is recorded but the study definition considers ‘slow responders’ as initial treatment failures, it can be inferred from (i) improvement of signs and symptoms between baseline and test-of-cure, and (ii) not requiring rescue treatment during initial treatment. Importantly, reflecting both routine clinical practice and a number of study definitions, detection of parasites at test-of-cure should not preclude the subsequent development of relapse, so long as points (i) and (ii) are met.

Relapse Where relapse is not directly recorded in the IPD as an efficacy outcome, the event can be inferred from two or more of: (i) the need for rescue treatment within 6 months of initial cure assessment (test-of-cure), (ii) the presence of a positive tissue aspirate, and (iii) in addition to a recurrence of compatible signs and symptoms.

with findings by Dahal et al, a significant proportion of contributing studies do not directly define relapse as a study outcome[90]. Instead, for most studies, a relapse event can be inferred from patients achieving initial cure who subsequently do not meet the definition of ‘definite cure’, which itself is typically defined as patients requiring rescue treatment.

Similar to the approach described for identifying patients that achieve initial cure, access to IPD allows relapse events to be inferred from other variables, including: definite cure status, tissue aspirates, timing of rescue treatment initiation, and patient signs and symptoms. Box 4.3 provides an IPD-based working definition of relapse.

Relapse is recorded and modelled as a binary outcome variable (occurred vs. not occurred). Unfortunately, modelling relapse as a time-to-event variable is not feasible, since *timing* information is (i) inconsistently presented across the contributed IPD, and (ii) where presented, is often limited to fixed, predetermined study visits (e.g. 3 months, 6 months).

4.3 Model Development

4.3.1 Sample Size

A common ‘rule of thumb’ is that at least 10–20 outcome events per predictor parameter (EPP) are needed to prevent model overfitting[226, 227]. However, both the validity of this threshold, and the broader premise that a single rule applies universally, have been increasingly debated in the prediction modelling literature[228].

A common ‘rule of thumb’ is that at least 10–20 outcome events per predictor parameter (EPP) are needed to prevent model overfitting[226, 227]. However, both the validity of this threshold, and the broader premise that a single rule applies universally, have been increasingly debated in the prediction modelling literature[228].

Responding to these concerns, Riley et al. (2018) proposed a principled sample size calculation framework for prediction model development[229], implemented here using the `pmsampsize` R package[230]. For binary or time-to-event outcomes, this approach defines minimum sample size requirements based on three criteria: (i) limited overfitting, operationalised as a global shrinkage factor ≥ 0.9 ; (ii) a small absolute difference (≤ 0.05) between apparent and adjusted Nagelkerke’s R^2 ; and (iii) precise estimation (within ± 0.05) of the average outcome risk. In the present study, where sample size and relapse rate are fixed, this framework was instead used to calculate the *maximum number of predictor parameters* supported by the data. In the absence of previously published relapse prediction models, a Nagelkerke R^2 of 0.15 was assumed, as recommended by the authors[229]. The maximum number of supported predictor parameters based on variations of criterion (i) are summarised in Table 4.1 for the ISC and EA datasets across their respective sample sizes and relapse rates.

Table 4.1: Maximum number of predictor parameters supported by the available IPD in the ISC and East Africa datasets. Calculations performed for a range of minimum permitted shrinkage factors (criterion 1) as per Riley et al’s methodology[229]. The remaining two criteria were unchanged from Riley et al’s recommendations (see text). For all calculations, the minimum permitted shrinkage factor was the criterion limiting the number supported predictor parameters. Fixed sample sizes — ISC: 4,599 patients, 228 relapses; East Africa: 2,051 patients, 99 relapses. Assuming Nagelkerke’s $R^2 = 0.15$.

Shrinkage factor	Maximum supported predictor parameters	
	ISC models	EA models
0.95	12	5
0.90 ¹	25	11
0.85	40	17
0.80	58	25

¹ Recommended by Riley et al[229].

ISC Model

With a total of 228 relapses identified in 4,599 participants (5.0% event rate), the maximum number of predictor parameters satisfying Riley et al.’s three criteria,

with a global shrinkage factor of at least 0.90, is 25 (corresponding to 8.86 EPP). Relaxing the permitted shrinkage factor from 0.90 to 0.85 allows for up to 40 predictor parameters (corresponding to 5.57 EPP).

East Africa Model

With a total of 99 relapses identified in 2,051 participants (4.8% event rate), the maximum number of predictor parameters satisfying Riley et al.'s three criteria, with a global shrinkage factor of at least 0.90, is 11 (corresponding to 8.81 EPP). Relaxing the permitted overall shrinkage from 0.90 to 0.85 allows for 17 predictor parameters, corresponding to 5.82 EPP.

4.3.2 Candidate Predictors

All model variables, including candidate predictors, are listed in Table 4.2.

For the ISC models, a total of 17 candidate predictor parameters are included (16 when excluding parasite grade), corresponding to a study-specific random intercept term and 12 (11) participant-level candidate predictors, EPP = 13.41 (14.25). Assuming a Nagelkerke's R² of 0.15, this EPP should comfortably satisfy sample size considerations.

For the East Africa models, a total of 14 candidate predictor parameters are included (13 when excluding parasite grade), corresponding to the same variables included in the ISC models minus treatment group, due to convergence issues discussed below, EPP = 7.07 (7.62). Assuming a Nagelkerke's R² of 0.15, this EPP should satisfy sample size calculations when the minimum permitted global shrinkage factor is reduced from 0.90 to 0.85.

Full specification of all candidate predictors are presented in Table 4.2.

The following points were considered when selecting candidate predictors:

- To avoid excessive missing data, predictors must be available for at least 50% of participants.
- Predictors should be routinely measured, or at least available for measurement, in the majority of treatment centres in endemic areas.
- Since the model is applied at the time of initial cure, predictors should be available at or *prior* to this time point (typically measured within a month of starting treatment).
- Predictors should be *preferentially* included if (i) they have previously been shown to predict relapse, or (ii) other compelling reasons exist to include — such as expert opinion and arguments supporting a causal association between the predictor and relapse. Discussions in Section 2.3 provide further insight.

Variable	Specification (categories)	dof	log	ISC		EA	
				PG	\overline{PG}	PG	\overline{PG}
Relapse (outcome)	Categorical (2)	-	-	✓	✓	✓	✓
Study	Random intercept	1	-	✓	✓	✓	✓
Treatment	Categorical (3)	2	-	✓	✓	✗	✗
Sex	Categorical (2)	1	-	✓	✓	✓	✓
Malnutrition	Categorical (3)	2	-	✓	✓	✓	✓
Anaemia severity	Categorical (2)	1	-	✓	✓	✓	✓
Age	Continuous ¹	3	No	✓	✓	✓	✓
Fever duration	Continuous	1	Yes	✓	✓	✓	✓
WBCs	Continuous	1	Yes	✓	✓	✓	✓
Spleen size	Continuous ²	1	No	✓	✓	✓	✓
Platelets	Continuous	1	Yes	✓	✓	✓	✓
ALT	Continuous	1	Yes	✓	✓	✓	✓
Creatinine	Continuous	1	Yes	✓	✓	✓	✓
Parasite grade ³	Continuous	1	No	✓	✗	✓	✗

¹ Age is modelled as a cubic polynomial.

² A small additive constant (+1) is added to spleen size to allow inclusion of patients with non-palpable spleens.

³ Parasite grade is recorded on a semi-quantitative logarithmic scale, ranging from 1+ to 6+.

Table 4.2: Variables used in the development of the Indian subcontinent (ISC) and East Africa (EA) models. Models are fitted either with parasite grade (PG) or without parasite grade (\overline{PG}). -: not applicable, ✓: included in model, ✗: not included in model, dof: degrees of freedom, log: whether modelled on a natural logarithmic scale, ALT: alanine aminotransferase, WBC: white blood cell.

- Sample size considerations explored in Section 4.3.1 should guide the maximum number of predictor parameters to prevent model overspecification and overfitting.
- To facilitate model convergence, excessive collinearity should be avoided.

When considering the final point, it is worth highlighting that model convergence issues can occur due to collinearity not only between predictors, but also between predictors and random-effects (clustering) structures. Given the significant heterogeneity in study design and outcome definitions, and further reasons addressed further in Section 4.3.5, each contributing study is modelled as a random-intercept term. Consequently, caution must be exercised with categorical predictors that are uniquely, or near-uniquely, identified at the study level.

Treatment

Treatment regimen — already established as an important predictor relapse — represents a categorical predictor affected by collinearity at the study level. For example, including a treatment category of *14 days of paromomycin* in the ISC model would lead to convergence failure when included in a model with study as a random-intercept term. As can be appreciated from Figure 5.2, only one contributing study (Sundar 2009) includes patients receiving this treatment regimen. Consequently, a model including both would not be able to distinguish between relapse risk related to the study or the treatment regimen.⁵

Considering the distribution of treatment regimens across studies within the ISC, treatment will be categorised into three categories that occur in at least three studies (see Table 5.1 and Figure 5.2):

- Single dose liposomal amphotericin B (10 mg/kg)
- 28 days miltefosine (standard dose)
- Other

As a consequence of creating a ‘catch-all’ treatment group: *Other*, some of the relationship between treatment and relapse not already accounted for at the study level will be lost. This is an important limitation and addressed further in the discussion (Chapter 7).

Unfortunately, convergence issues preclude the inclusion of treatment as a categorical predictor in the East Africa models. This is not surprising, given fewer participants, fewer studies, and higher treatment-study collinearity when compared to the ISC models (see Figure 6.2). Convergence issues persisted despite exploring different treatment groupings, including separate groups for SSG and SSG/PM combination therapy. Therefore, in the East Africa models, the impact of treatment on relapse risk is incorporated into the study level random intercept.

Malnutrition

Malnutrition is a well-established determinant of progression from asymptomatic infection to clinical VL and of adverse outcomes following initial treatment, including treatment failure and mortality[1, 201]. Evidence linking malnutrition to relapse, however, remains sparse. This may partly reflect the absence of a unified

⁵Or more technically, this induces near-non-identifiability between the fixed treatment effect and the study-level random intercept, resulting in an ill-conditioned information matrix and failure of numerical optimization.

Table 4.3: Malnutrition severity definitions by age group. BMI: body-mass-index, WFH: weight-for-height, yrs: years, [,]: including range limit, (,): not including range limit

Age group	Metric	Severe	Moderate	Mild/normal
[0, 5) yrs	WFH z-score	($-\infty$, -3]	(-3, -2]	(-2, ∞)
[5, 19) yrs	BMI z-score	($-\infty$, -3]	(-3, -2]	(-2, ∞)
[19, ∞) yrs	BMI (kg/m^2)	(0, 16)	[16.0, 17.0)	[17, ∞)

framework for defining malnutrition across age groups, leading to frequent omission or inconsistent classification in prognostic factor and prediction model studies. Anthropometric assessment of malnutrition is inherently age-specific, and no formal guidance exists on how to combine measures across the life course.

We therefore adopt a pragmatic three-level severity scale using age-appropriate metrics (Table 4.3). In children under five years, malnutrition is classified using weight-for-height (WFH) z-scores, consistent with WHO guidelines on the definition of acute malnutrition[231]. For individuals aged five to under nineteen years, BMI-for-age z-scores (BMI z-scores) are used with identical cut-points, supported by the WHO 2007 growth reference, which was explicitly constructed to ensure continuity with the under-five standards at age five[232]. In adults aged 19 years and over, malnutrition severity is defined using established BMI thresholds. Cole et al. showed that adult BMI cut-offs of 16 and 17 kg/m^2 at 18 years correspond approximately to BMI-for-age z-scores of -3 and -2 in children and adolescents, providing a statistical basis for approximate continuity of severity definitions across the adolescent-adult boundary[233]. While imperfect, this approach preserves broadly comparable degrees of nutritional deficit across age groups.

For modelling purposes, individuals with mild ($18.5 \leq \text{BMI} < 25 \text{ kg}/\text{m}^2$) or obesity ($\text{BMI} \geq 25 \text{ kg}/\text{m}^2$), or equivalent z-scores ($z > -1$), were combined into a single *mild/normal* category due to small numbers.

The approach adopted here is consistent with methods used in previous studies of VL outcomes that also use age-specific anthropometric indicators to define pragmatic malnutrition severity groupings[96, 112, 125, 132]. z-scores were calculated using the `anthro` and `anthroplus` R packages for children < 5 years and ≥ 5 years, respectively[234, 235].

Anaemia

Anaemia was grouped into two categories: *severe* and *non-severe*, using haemoglobin cut-offs stratified by age and sex thresholds as defined by 2024 WHO guidelines[236]. Additional subdivision of the non-severe anaemia group was limited by the small number of participants in the mild and normal categories.

Age

To account for an anticipated non-linear relationship between age and relapse, age was modelled as a third-degree polynomial term (including linear, squared, and cubic components).

Parasite Grade

Baseline parasite grade, when available, was assessed from splenic, bone marrow, or lymph node aspirates. When reported, the logarithmic counting method of Chulay and Bryceson (1983) was either described or directly cited Additional file 1[52].

Logarithmic Transformations

A number of continuous predictors — including fever duration, spleen length, and all blood tests — were log-transformed to reduce skewness and thus stabilise variance and better approximate normality. For spleen size, +1 was added prior to log-transformation to accommodate zero values and avoid undefined logarithmic results. While recognising that this transformation implicitly assumes that non-palpable spleens can be modelled on the same continuous scale with palpable spleens, the approach was considered a pragmatic solution to avoid model overspecification.

All models assume that the adjusted log(odds) of relapse varies *linearly* with each continuous predictor when transformed as per Table 4.2. These assumptions are evaluated prior to model fitting through visual assessment of the univariable associations (as described below in Section 4.3.3, and presented in Figure C.17 for the ISC and Figure D.17 for East Africa) and, following model fitting, through inspection of the calibration plots.

4.3.3 Descriptive Analysis

All candidate predictors (marginal and study-specific) are summarised in both tabular and graphical forms.

The correlation between continuous predictors is illustrated with faceted scatter plots using the `ggpairs()` function from the `GGally` R package[237]. The correlation between continuous and categorical predictors is presented with faceted box-and-whisker plots.

Unadjusted relationships between the candidate predictors and the outcome (relapse) are presented in graphical form with 95% confidence intervals. Since this stage of the analysis is exploratory and descriptive, p-values are not presented to

avoid multiple hypothesis testing and inflation of type I error, with emphasis instead placed on the shape, magnitude, and uncertainty of observed associations[238].

For continuous predictors (excluding parasite grade), relapse is modelled using a generalised additive model (GAM) with a binomial error distribution and logit link, fitted using the `gam()` function in the R package `mgcv`[239]. The effect of the continuous predictor is represented by a smooth function estimated using penalised thin-plate regression splines. Model fitting is performed by penalised maximum likelihood, with the degree of smoothness selected automatically via restricted maximum likelihood (REML). Both relapse % and log(odds) of relapse are presented, allowing for both direct inspection of the relapse risk and visual assessment of the linearity assumption between the log(odds) of relapse and transformed continuous predictors.

For categorical predictors and parasite grade, relapse is presented using bar charts with 95% confidence intervals calculated using the Wilson method[240].

4.3.4 Missing Data

Data can be missing entirely from a study (systematic missingness), or affecting only certain patients within a study (sporadic missingness). Three principal sources of missing data were identified across both the ISC and East Africa harmonised datasets:

- planned non-capture at the study or site level, often resulting in systematic missingness,
- unplanned incomplete capture of predictors at the study level, resulting in sporadic missingness, and
- incomplete data retrieval, likely resulting in both systematic and sporadic missingness.

Patterns and extent of missingness in candidate predictors were explored and summarised using both tabular and graphical approaches. Study-stratified density plots were used to visualise missingness patterns stratified by study and are presented in the results chapters (Figures 5.7 and 6.7 presented in the Results chapters).

Multiple imputation with chained equations (MICE)⁶ was used to generate multiple imputed datasets[241, 242]. MICE was chosen for its flexibility in accommodating different variable types (continuous, binary, categorical, and count), its suitability for complex data structures including multilevel data, and the availability

⁶also known as fully conditional specification

Table 4.4: Multiple imputation with chained equations (MICE) methods used for univariable imputation of variables with missing data. ALT: alanine aminotransferase; BMI: body mass index; Crt: creatinine; Hb: haemoglobin; ISC: Indian subcontinent; n/a: not applicable; Plt: platelets; WBC: white blood cell count; WFH: weight-for-height.

Method	Variables	Description	R package
2l.lmer	age, weight, height, fever duration, all bloods (Hb, ALT, WBC, Plt, Crt), spleen size	Two-level normal model using <code>lme4::lme4()</code>	<code>mice</code> [243] <code>micemd</code> [246]
2l.zip	parasite grade (East Africa only)	Two-level zero-inflated Poisson model using Bayesian estimation	<code>countimp</code> [247]
2l.poisson	parasite grade (ISC only)	Two-level Poisson model using Bayesian estimation	<code>countimp</code> [247]
Passive imputation	BMI, BMI-for-age z-score, WFH z-score, age ² , age ³	Calculated during imputation from age, weight and height.	<code>mice</code> [243]
n/a	outcome, treatment (ISC only), sex, study (random intercept term)	No imputation performed (no missing data)	n/a

of well-established software implementations in R, notably the `mice` package and its extensions[243, 244].

Imputation was performed under the missing at random (MAR) assumption, whereby the probability of missingness depends only on observed data[245]. Congenerality between the imputation and analysis models was promoted by (1) including in the imputation model all candidate predictors, the outcome, and variables used to derive predictors (including haemoglobin, height, and weight), and (2) accounting for between-study heterogeneity using a multilevel imputation framework with study included as a random intercept[241]. The imputation methods used are summarised in Table 4.4.

The imputation predictor matrix was initially specified to be saturated and include all variables listed in Table 4.4 as well as variables without missing data

(outcome, treatment, and sex), and derived predictors (age polynomial terms, BMI, and anthropometric z-scores). Study was included as a random intercept. Derived predictors were passively imputed from age, weight, and height at the end of each iteration, and, to avoid circularity in the imputation model, were excluded as independent variables when imputing age, weight, and height.

Specification of the imputation model was an iterative process. Candidate imputation models were evaluated with respect to (1) convergence of the chained equations and (2) the extent to which the distributions of imputed values aligned with the corresponding observed data distributions. Model specification and selection were informed by visual inspection of diagnostic plots, including trace plots of the mean and standard deviation of imputed values across iterations, density plots comparing observed and imputed distributions, and scatter plots comparing observed and imputed data for key anthropometric relationships (weight versus height, weight versus age, and height versus age). Diagnostic plots are included in the [Supplementary Material](#). On visual inspection of the trace plots, twenty iterations per imputation were considered more sufficient to reach stable convergence.

Parasite grade was modelled as count data. One unit was subtracted from the original grading scale (1+ to 6+) to permit zero values. The final choice of model was guided by matching the distribution of parasite grades from the imputation model with the observed distribution in the non-missing data. In the ISC dataset, a Poisson model provided the best fit to the observed data. In contrast, for the East Africa dataset, a zero-inflated Poisson model better captured the excess of 1+ grades.

For each model development dataset, 30 imputations were generated with 20 iterations per imputation. This represented a balance between computational feasibility and the need to stabilise parameter estimates. As a general rule, it has been stated that the number of imputations should at least match the percentage of incomplete cases^[241]. For datasets with many variables, however, van Buuren suggests that this requirement can be relaxed such that the minimum number of imputations approximates the average missing data rate^[243]. The average missing data rates across candidate predictors were 14.3% for the ISC dataset and 14.9% for the East Africa dataset, whilst the proportion of cases with any missing data was over 80%. Since 80 iterations was considered too computationally intensive, a compromise of 30 imputations was chosen.

Unless specified otherwise, subsequent analyses were performed separately within each imputed dataset, with results pooled across imputations using Rubin's Rules^[245].

Grouping of anaemia severity and malnutrition status was performed after imputation, as described in Section [4.3.2](#).

4.3.5 Model Specification

Relapse was modelled using a multivariable generalised linear mixed-effects model (GLMM) with a logit link function. Predictors were transformed as previously described. Following the aforementioned transformations, continuous predictors were centred and scaled by their means and standard deviations, respectively, to improve model stability. Anticipated between-study heterogeneity was accounted for by including study as a random intercept term[248]. While introducing significant methodological complexity, a number of benefits gained by accounting for between-study heterogeneity[212]:

- Ignoring clustering can result in relapse probability estimates that are biased towards the marginal population estimate[249].
- Variation in model performance measures can be compared and contrasted across the included studies, allowing insights into sources of heterogeneity[250].
- Allows for improved generalisability of the model to new settings and populations[212, 250].

GLMMs were fitted using the `glmer()` function from the `lme4` R package[251]. This approach estimates fixed effects while accounting for random effects by maximising the marginal likelihood, which integrates over the random-effects distribution. For GLMMs, this integration cannot be solved exactly and is approximated using the Laplace approximation for binomial outcomes. Model parameters were estimated using maximum likelihood. To enhance model stability and improve convergence, optimisation was performed using the `nloptwrap` optimizer with an increased number of iterations and strict convergence criteria (maximum evaluations = 200,000; absolute tolerances for the objective function and parameters = 1×10^{-8}).

The saturated ISC model (with parasite grade and treatment groups included) is presented in Box 4.4. For the East Africa models, the treatment group predictors are omitted due to collinearity at the study level, as previously discussed. For both ISC and East Africa, models are also fitted without parasite grade, given its frequent absence from routine clinical practice.

Box 4.4: Saturated model specification

$$\log \left(\frac{\Pr(Y_{ij} = 1)}{\Pr(Y_{ij} = 0)} \right) = \beta_0 + \beta_{\text{sex}} \cdot \text{sex}_{ij} + \beta_{\text{age}} \cdot \text{age}_{ij} + \beta_{\text{age}^2} \cdot \text{age}_{ij}^2 + \beta_{\text{age}^3} \cdot \text{age}_{ij}^3 + \beta_{\text{treat-sda}} \cdot \text{treat-sda}_{ij} + \beta_{\text{treat-oth}} \cdot \text{treat-oth}_{ij} + \beta_{\text{mal-mod}} \cdot \text{mal-mod}_{ij} + \beta_{\text{mal-sev}} \cdot \text{mal-sev}_{ij} + \beta_{\text{anaemia-sev}} \cdot \text{anaemia-sev}_{ij} + \beta_{\text{fever-dur}} \cdot \text{fever-dur}_{ij} + \beta_{\text{spleen-cm}} \cdot \text{spleen}_{ij} + \beta_{\text{para}} \cdot \text{para}_{ij} + \beta_{\text{wbc}} \cdot \text{wbc}_{ij} + \beta_{\text{plat}} \cdot \text{plat}_{ij} + \beta_{\text{alt}} \cdot \text{alt}_{ij} + \beta_{\text{creat}} \cdot \text{creat}_{ij} + \mu_j.$$

Where: i represents the i th participant in study j , Y is the relapse outcome (1 = relapse, 0 = no relapse), and

$$Y_{ij} \sim \text{Bernoulli}(E[Y_{ij}]) \\ \mu_j \sim \mathcal{N}(0, \tau^2).$$

β_0 is the fixed model intercept, and β_k are the fixed effect coefficients for each candidate predictor k , corresponding to sex (1 = male, 0 = female), age, treatment group (reference = standard dose miltefosine, **treat-sda**: 10 mg/kg single dose liposomal amphotericin B, **treat-oth**: other), malnutrition severity (reference = mild/normal, **mal-mod**: moderate, **mal-sev**: severe), anaemia severity (1 = severe, 0 = non-severe), fever duration, spleen size, **para**: parasite grade, **wbc**: white blood cell count, **plat**: platelet count, **alt**: alanine transaminase, and **creat**: creatinine. The random intercept μ_j captures between-study heterogeneity, with variance τ^2 . All continuous candidate predictors, except age, spleen size, and parasite grade, are log-transformed, centred by their mean, and scaled by their standard deviation. Age is centred and scaled, and modelled as a third-degree polynomial term. Spleen size is log-transformed after adding +1.

4.3.6 Variable Selection

The final predictor set was determined using backwards variable selection with cutoff $p < 0.10$. Univariable selection was avoided to prevent exclusion of predictors that may demonstrate significance only in the presence of other predictors, and to reduce the risk of overfitting[173].

A variety of methods have been explored for performing variable selection with multiple imputed datasets[252, 253]. In accordance with recommendations by Austin et al. and Wood et al., Rubin's Rules were used to combine p-values for predictor significance at each selection stage across the imputed datasets. At each stage, the full model is fitted in each of the 30 imputed datasets and the estimated regression coefficients and their standard errors are pooled using Rubin's Rules. The candidate predictor with the highest pooled p-value above the 0.10 threshold was removed,

and the process repeated until all remaining predictors had pooled p-values below 0.10[245, 252, 253]. For categorical predictors with over two groups (malnutrition, treatment), predictor significance was assessed with the D1 multivariate Wald test as implemented in the `mice` R package[243, 254]. For age, lower-order polynomial terms were retained in the model, whilst higher-order terms remained.

Variable selection was performed in a bespoke R script, available in the [Supplementary Material](#).

4.3.7 Model Performance

In accordance with the TRIPOD-Cluster reporting guidelines, measures of calibration and discrimination are presented for the prediction models overall and for each study (cluster) separately[212]. Wynants et al. address a number of decisions that must be made when evaluating prediction model performance in clustered data[249]. As described by Wynants et al, within-study (conditional) performance measures are presented reflecting model performance when evaluated at the study level[249, 255]. Assessment of model performance heterogeneity across studies is presented with forest plots using a combination of fixed and random effects meta-analyses[174, 255].

Discrimination

Model discrimination was assessed using the c (concordance)-statistic. For models with binary outcomes, the c-statistic is equivalent to the AUC, and is the probability that a randomly selected case (patient who experiences a relapse event) has a higher predicted relapse probability than a randomly selected non-case (patient who does not experience a relapse event).

Study-specific c-statistics were estimated for each imputed dataset using the R package: `pROC`[256]. Standard errors and 95% confidence intervals were calculated using the bootstrap method with 2000 resamples.

As suggested by Burgess et al., study-specific c-statistics were first pooled across imputed datasets using Rubin's Rules *prior* to estimating the summary (overall) c-statistic using meta-analysis[257]. As recommended by van Klaveren et al.[255] and Debray et al.[174], overall c-statistics are estimated using a random-effects meta-analysis approach. Between-study variance (τ^2) is estimated using restricted maximum likelihood (REML), and variance of the pooled c-statistic is estimated with the Hartung-Knapp-Sidik-Jonkman method[258].

For comparison, a fixed-effects meta-analysis approach to estimating the overall c-statistic is also presented, weighted by the number of event/non-event pairs in

each study. This estimate is equivalent to the c-statistic obtained from comparing pairs of events and non-events within the same study only[249, 259].

Given the often small number of relapses in individual studies, estimates of study-specific c-statistic uncertainty (variance, confidence intervals) are likely to be imprecise and should be interpreted cautiously. To reduce potential bias introduced from small sample sizes, only studies with >5 events are included in the meta-analyses[260, 261].

Calibration

Model calibration assesses agreement between predicted and observed relapse probability, and was evaluated with calibration plots, calibration slope, and calibration intercept (calibration-in-the-large, CITL). Calibration slope and intercept are estimated conditional on the study, and represent within-study performance measures[249].

Calibration intercept describes whether a model's average predicted risk differs from the overall observed event rate. A positive intercept indicates that the model systematically underestimates relapse risk, whilst a negative intercept indicates systematic overestimation. The calibration intercept is estimated by fitting a logistic GLMM with relapse as the outcome, study as a random intercept, and the linear predictor from the original model as an offset term. The estimated fixed intercept then corresponds to the within-study calibration intercept, with an expected value of zero when assessed in the development dataset (apparent performance)[249].

Calibration slope assesses whether a model's predicted risks are systematically too extreme or too conservative. A slope of 1 indicates perfect calibration and is the expected value when evaluating apparent performance in the development dataset. Values less than 1 indicate overfitting (overly extreme predictions), and values greater than 1 indicate underfitting (overly moderate predictions). It was estimated by fitting a further logistic GLMM with relapse as the outcome, study as a random intercept, and the linear predictor from the original model included as both a fixed and random effect (i.e. a random slope term)[249]. In this formulation, the fixed effect of the linear predictor represents the within-study calibration slope. To mitigate model convergence issues arising from the inclusion of studies with small sample sizes, we adopted a Bayesian approach. Parameters were estimated using the Hamiltonian Monte Carlo algorithm as implemented in Stan, accessed via the R package `brms`[262, 263]. Posterior sampling was conducted using four chains with 2,500 iterations per chain (including warm-up), and the

target acceptance rate was increased (`adapt_delta = 0.95`) to reduce divergent transitions. Default weak priors were used.

Heterogeneity in calibration slope and intercept across studies is summarised using a the same random effects approach to that described for the c-statistic. Study-specific estimates were initially calculated within each imputed dataset prior to pooling across imputations using Rubin's Rules[257]. Summary measures were then estimated using random-effects meta-analysis with REML estimation of between-study variance and Hartung-Knapp-Sidik-Jonkman confidence intervals[258].

Calibration plots allow direct visual assessment of the relationship between observed (y -axis) and predicted (x -axis) relapse probabilities. Firstly, all imputed datasets were stacked into a single dataset. Predicted relapse probabilities were then calibrated to each study using logistic recalibration of the intercept and divided into deciles. Observed probabilities were calculated for each decile and 95% confidence intervals estimated using Wilson's method[240]. The smoothed relationship between observed and predicted relapse probabilities was estimated using a generalised additive model, as previously described. To account for the stacking of 30 imputed datasets, all standard errors were inflated by $\sqrt{30}$.

Following the same methodology for standard calibration plots, calibration was also illustrated for selected predictor subgroups by plotting both observed and predicted probabilities on the y -axis against predictor subgroups on the x -axis.

4.4 Internal Validation

Internal validation is performed using bootstrap resampling with 500 bootstrap samples, as recommended by Collins et al[196]. Bootstrapping evaluates all aspects of the model building process to quantify and adjust for overfitting and optimism in model performance[196]. Bootstrapping is performed at the overall patient, which as described by Bouwmeester et al., provides accurate optimism estimates in clustered data[264].

For each of the 500 bootstrap samples of the original dataset, all aforementioned model development steps, including multiple imputation and variable selection, were undertaken (see Figure 4.1). For each of the resulting 500 bootstrap models, pooled performance measures were evaluated across (i) the imputed bootstrap datasets used to derive the bootstrap model and (ii) the imputed original datasets used to derive the original model. The mean pooled performance differences between (i) and (ii) were then subtracted from the original model apparent performance measures to obtain the optimism-adjusted performance measures. Uniform shrinkage

of the parameter estimates was performed by multiplying the fitted regression coefficients by the optimism-adjusted calibration slope and re-estimating the model intercept with logistic recalibration.

Given the computational intensity of the model development process (4 hours per model development), bootstrap validation was implemented using a high performance computing cluster based in the Nuffield Department of Medicine at the University of Oxford.⁷

4.5 Summary

In summary, this chapter has described the methods used to develop and internally validate four prognostic models for six-month VL relapse (HIV-negative patients) using IPD from the IDDO VL platform: two models for the ISC (with and without baseline parasite grade) and two analogous models for East Africa.

Contributed datasets were standardised to the CDISC-SDTM format, converted to a single analysis-ready dataset, and filtered using pre-specified study- and participant-level eligibility criteria that ensure the population at risk comprises patients who achieved an initial cure. Candidate predictors were selected based on clinical relevance and data availability, and were transformed as appropriate (e.g. logarithmic or polynomial transformations). Predictor specification was informed by considerations of collinearity, including grouping of treatment variables to reduce study-level dependence and support model convergence. Sample-size considerations followed the framework of Riley et al. to guide the maximum allowable degrees of freedom.

Modelling used multivariable GLMMs with study as a random intercept, fitted on multiply imputed datasets (MICE, 30 imputations, multilevel imputation with study random intercept and diagnostics to check plausibility). Backwards selection (pooled p-values, cutoff $p < 0.10$) determined the final predictor set, with coefficients and uncertainty combined with Rubin's rules at each variable selection stage. Internal validation applied full bootstrap resampling ($b = 500$), repeating multiple imputation and selection within each bootstrap to quantify optimism. The final models were uniformly shrunk by the optimism-adjusted calibration slope.

⁷<https://www.medsci.ox.ac.uk/for-staff/resources/bmrc>

5

Results: Indian Subcontinent Models

Contents

5.1 Descriptive Analysis	76
5.1.1 Study Characteristics	76
5.1.2 Patient Characteristics	84
5.1.3 Univariable Associations	90
5.1.4 Missing Data	92
5.2 Model Results	93
5.2.1 Model Specification and Coefficient Estimates	93
5.2.2 Model Performance and Internal Validation	97
5.3 Summary	99

5.1 Descriptive Analysis

On application of the eligibility criteria to the IDDO VL data platform, a total of 19 studies and 4,599 patients were selected for inclusion[64, 67, 71, 94, 265–279]. At the participant selection stage, 27 participants (0.6%, 27/4860) were excluded due to having a positive HIV test, and 201 (4.1%, 201/4860) were excluded due to not achieving initial cure. A flow diagram is presented in Figure 5.1.

5.1.1 Study Characteristics

Key characteristics of the 19 included studies are presented in Table 5.1.

A median of 144 patients were recruited in each study, ranging from 6 to 928 patients (IQR: 87 to 330 patients). Patients were recruited between 2000 and 2017

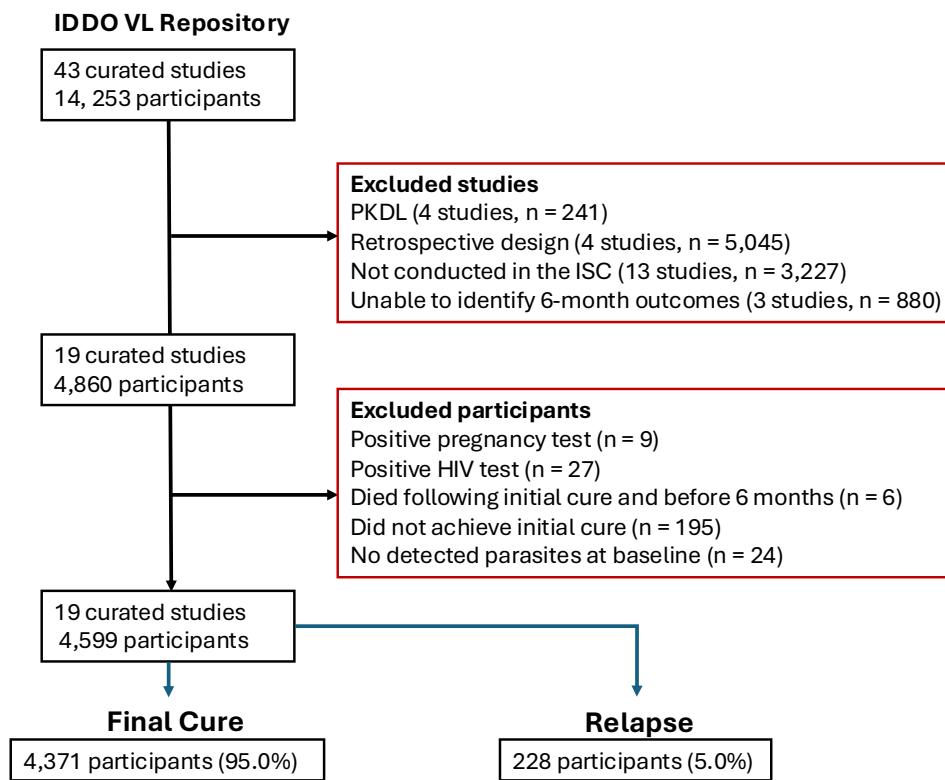


Figure 5.1: Flow diagram showing the studies and patients excluded from Indian subcontinent model development. HIV: human immunodeficiency virus; IDDO: Infectious Diseases Data Observatory; ISC: Indian subcontinent; PKDL: post kala-azar dermal leishmaniasis; VL: visceral leishmaniasis.

from sites in India (15 studies, 4,171 [90.7%] patients)[64, 67, 71, 94, 265, 266, 268, 269, 273–279] and Nepal (4 studies, 428 [9.3%] patients)[267, 270–272]. All Indian studies recruited patients from the state of Bihar in northeastern India, with the majority of studies led by investigators from either a government referral hospital in Patna (Rajendra Memorial Research Institute of Medical Sciences: RMRIMS) (4 studies, 1,125 [24.5%] patients)[64, 266, 268, 269], or a private research facility in Muzaffarpur (Kala-azar Medical Research Centre: KAMRC) (9 studies, 2,652 [57.7%] patients)[67, 94, 265, 273–277, 279]. All Nepalese studies were led by investigators based at the B.P. Koirala Institute of Health Sciences, a health sciences university in the city of Dharan, Koshi Province, eastern Nepal.

Table 5.1: Key characteristics of included studies, ordered by lead author and year of publication/protocol. Where information not presented in the publication, information is extracted from the study protocol. Total patient number, number of relapses, and treatment arms reflect information provided in the contributed datasets. -: not reported; ABLE: Amphotericin B lipid emulsion (Bharat Serum and Vaccines Ltd.); alt.: alternative; ABD: Amphotericin B deoxycholate; BD: twice daily (bis die); BMGF: Bill and Melinda Gates Foundation; BPKIHS: B.P. Koirala Institute of Health Sciences; D: Day(s); EC: European Commission; Govt.: Government; ICMR: Indian Council of Medical Research; IM: Intramuscular; KAMRC: Kala-azar Medical Research Center; LAMB: Liposomal amphotericin B (Gilead formulation unless otherwise specified); MF: Miltefosine; mg/kg: milligrams per kilogram; NCVBDC: National Center for Vector Borne Diseases Control; OD: once daily (omni die); P: Publication; PATH: Program for Appropriate Technology in Health; PO: per os (oral); PM: Paromomycin; Pr: Protocol; Ref: Reference; RMRIMS: Rajendra Memorial Research Institute of Medical Sciences; SSG: Sodium stibogluconate; TDR: UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases; UN: United Nations; WHO: World Health Organization

Study ¹	Title	Journal	Sponsor/funding	Location(s)	Study design	Study arm(s)	Age (years)	Study period	n (model)	Relapses (%)
Bhattacharya 2007[64]	Phase 4 trial of miltefosine for the treatment of Indian visceral leishmaniasis	J Infect Dis	ICMR	13 locations in Bihar, India (outpatient setting)	Open label; phase 4; safety/efficacy	(1) MF, PO, 28 D, dosed according to age and weight ^{2,3}	2–65	2002–2004	352 ⁴	22 (6.3)
Chakraborty 2008[265]	Human placental extract offers protection against experimental visceral leishmaniasis: a pilot study for a phase-I clinical trial	Am J Trop Med Hyg	Indian Council of Scientific and Industrial Research; Albert David Ltd.	KAMRC, Muzaffarpur, India	“Pre-phase 1”; pilot/preliminary	(1) Human placental extract, 2.06 mg, IM, single dose; (2) ABD, 1 mg/kg, IV, alt. days for 30 D ⁵	≥ 5	2003–2005	6	0 (0)
Das 2009[266]	A controlled, randomized nonblinded clinical trial to assess the efficacy of amphotericin B deoxycholate as compared to pentamidine for the treatment of antimony unresponsive visceral leishmaniasis cases in Bihar, India	Ther Clin Risk Manag	-	RMRIMS, Patna, India	Randomised; open label; efficacy	(1) ABD, 1 mg/kg, IV, alt. days for 30 D; (2) pentamidine, 4 mg/kg, IM, alt. days for 30 D	6–60	2002	73	5 (6.8)
Koirala 2003[267]	Phase IV trial of miltefosine in the treatment of visceral leishmaniasis	Protocol only	-	BPKIHS, Dharan, Nepal	Open label; phase 4; safety/efficacy	(1) MF, PO, 28 D, dosed according to age and weight ^{2,3}	2–65	2003–2004	116	12 (10.3)
Pandey 2016[268]	Pharmacovigilance of miltefosine in treatment of visceral leishmaniasis in endemic areas of Bihar	Am J Trop Med Hyg	NCVBDC, Govt. of India; World Bank	4 locations in Bihar, India	Open label; safety/efficacy	(1) MF, PO, 28 D, dosed according to age and weight ^{2,6}	6–70	2012–2015	600	45 (7.5)
Pandey 2017[269]	Efficacy and safety of liposomal amphotericin B for visceral leishmaniasis in children and adolescents at a tertiary care center in Bihar, India	Am J Trop Med Hyg	-	RMRIMS, Patna, India	Open label; safety/efficacy	(1) LAMB, 10 mg/kg, IV, single dose	<15	2014–2016	100	2 (2.0)

continued on next page

Table 5.1: continued

Study ¹	Title	Journal	Sponsor/funding	Location(s)	Study design	Study arm(s)	Age (years)	Study period	n (model)	Relapses (%)
Rijal 2003[270]	Treatment of visceral leishmaniasis in south-eastern Nepal: decreasing efficacy of sodium stibogluconate and need for a policy to limit further decline	Trans R Soc Trop Med Hyg	WHO, Geneva University Hospital; Novartis Foundation	BPKIHS, Dharan, Nepal	Non-randomised; efficacy	SSG, 20 mg/kg, IV/IM ⁷ OD, 30 D, extended to 40 D if positive aspirate at 30 D. Either (1) in hospital or (2) first 5–7 days in hospital;	All	2000–2001	102	1 (1.0)
Rijal 2010(A)[271]	Efficacy and safety of liposomal amphotericin B in Nepalese patients with visceral leishmaniasis	Protocol only	TDR	BPKIHS, Dharan, Nepal	Phase 2/3; safety/efficacy	(1) LAMB, 3 mg/kg, IV, OD, D1–5 (21 mg/kg total dose)	12–65	2010–2011	32	1 (3.1)
Rijal 2010(B)[272]	Clinical risk factors for therapeutic failure in kala–azar patients treated with pentavalent antimonials in Nepal	Trans R Soc Trop Med Hyg	EC (5th Framework Programme)	BPKIHS, Dharan, Nepal	Prospective cohort	(1) SSG, 20 mg/kg, IV/IM ⁷ OD, 30 D	-	2001–2003	178	1 (0.6)
Sundar 2007[71, 280]	Injectable paromomycin for visceral leishmaniasis in India	N Engl J Med	PATH (including UN, BMGF, TDR)	4 locations in Bihar, India	Randomised; open label; phase 3; non-inferiority; safety/efficacy	(1) PM, 15 mg/kg, IM, OD, 21 D; (2) ABD, 1 mg/kg, IV, alt. days for 30 D	5–55	2003–2004	250 ⁴	4 (1.6)
Sundar 2008(A)[273, 281]	New treatment approach in Indian visceral leishmaniasis: single-dose liposomal amphotericin B followed by short-course oral miltefosine	Clin Infect Dis	Banaras Hindu University	KAMRC, Muzaffarpur, India	Partially randomised; open label; phase 2; non-comparative; sequential; triangular	LAMB, 5 mg/kg, IV, single dose, either (1) alone, or (2,3,4) in combination with MF, 50 mg, PO, BD, D2–9, D2–11, or D2–15; and (5) combination of LAMB, 3.75 mg/kg, IV, single dose, and MF, 50 mg, PO, BD, D2–15	≥ 12	2006–2007	225	7 (3.1)
Sundar 2008(B)[274]	Safety of a pre-formulated amphotericin B lipid emulsion for the treatment of Indian kala–azar	Trop Med Int Health	Bharat Serum and Vaccines Ltd.	KAMRC, Muzaffarpur, India	Non-randomised; non-comparative; open label; phase 2; safety/efficacy	ABLE, IV, OD, 3 D at (1) 5 mg/kg (2) 4 mg/kg, or (3) 3 mg/kg	12–65	2004–2005	45	4 (8.9)
Sundar 2009[275, 282]	Short-course paromomycin treatment of visceral leishmaniasis in India: 14-day vs 21-day treatment	Clin Infect Dis	Banaras Hindu University	KAMRC, Muzaffarpur, India	Randomised; open label; phase 3; safety/efficacy	PM, 15 mg/kg, IM, OD, either (1) 14 D or (2) 21 D	5–55	2007–2008	307	26 (8.5)
Sundar 2010[276, 283]	Single-dose liposomal amphotericin B for visceral leishmaniasis in India	N Engl J Med	Banaras Hindu University	KAMRC, Muzaffarpur, India	Randomised; open label; phase 3; safety/efficacy	(1) ABD, 1 mg/kg, IV, alt. days for 30 D; (2) LAMB, 10 mg/kg, IV, single dose	2–65	2008–2009	412	14 (3.4)

continued on next page

Table 5.1: continued

Study ¹	Title	Journal	Sponsor/funding	Location(s)	Study design	Study arm(s)	Age (years)	Study period	n (model)	Relapses (%)
Sundar 2011[277, 284]	Ambisome plus miltefosine for Indian patients with kala-azar	Trans R Soc Trop Med Hyg	Banaras Hindu University	KAMRC, Muzaffarpur; RMRIMS, Patna, India	Open label; phase 2; safety/efficacy	(1) LAMB, 5 mg/kg, IV, single dose, with MF, 2.5 mg/kg/day, PO, D2-D15 ⁸	2–65	2007–2009	128	5 (3.9)
Sundar 2012[67]	Efficacy of miltefosine in the treatment of visceral leishmaniasis in India after a decade of use	Clin Infect Dis	EC (Kaladrug-R); Sitaram Memorial Trust	KAMRC, Muzaffarpur, India	Open label; safety/efficacy	(1) MF, PO, 28 D, dosed according to age and weight ^{2, 3}	6–70	2009–2010	571	34 (6.0)
Sundar 2014[278, 285]	Efficacy and safety of amphotericin B emulsion versus liposomal formulation in Indian patients with visceral leishmaniasis: a randomised, open-label study	PLoS Negl Trop Dis	Bharat Serums and Vaccines Ltd; Department of Science and Technology, Govt. of India	4 locations in Bihar, India.	Randomised; open label; phase 3; safety/efficacy	(1) LAMB, 15 mg/kg, IV, single dose; (2) ABLE, 15 mg/kg, IV, single dose	5–65	2009–2011	144 ⁴	9 (6.3)
Sundar 2015[279, 286]	Single-dose indigenous liposomal amphotericin B in the treatment of Indian visceral leishmaniasis: A phase 2 study	Am J Trop Med Hyg	Lifecare Innovations; Department of Science and Technology, Govt. of India	KAMRC, Muzaffarpur, India	Non-randomised; non-comparative; open label; phase 2; safety/efficacy	LAMB (Lifecare Innovations), IV, single dose (1) 10 mg/kg or (2) 15 mg/kg	12–60	2012–2013	30	3 (10.0)
Sundar 2019[94, 287]	Effectiveness of single-dose liposomal amphotericin B in visceral leishmaniasis in Bihar	Am J Trop Med Hyg	Banaras Hindu University	KAMRC, Muzaffarpur, India	Observational; efficacy	(1) LAMB, 10 mg/kg, IV, single dose	All	2013–2017	928	33 (3.6)

¹ Study name is composed of the lead author and year of publication or most recent protocol.² Miltefosine dosing (i) ≥ 12 years and ≥ 25 kg: 50 mg BD; (ii) ≥ 12 years and <25 kg: 50 mg OD; (iii) <12 years and <25 kg: 2.5 mg/kg/day.³ Dosing in <12 years and <25 kg: in divided doses⁴ Contributed number of participants from KAMRC site only.⁵ ABD arm not presented in publication.⁶ Dosing in <12 years and <25 kg: not further described.⁷ Dosing not further described.⁸ SSG route of administration (IV vs IM) not specified.

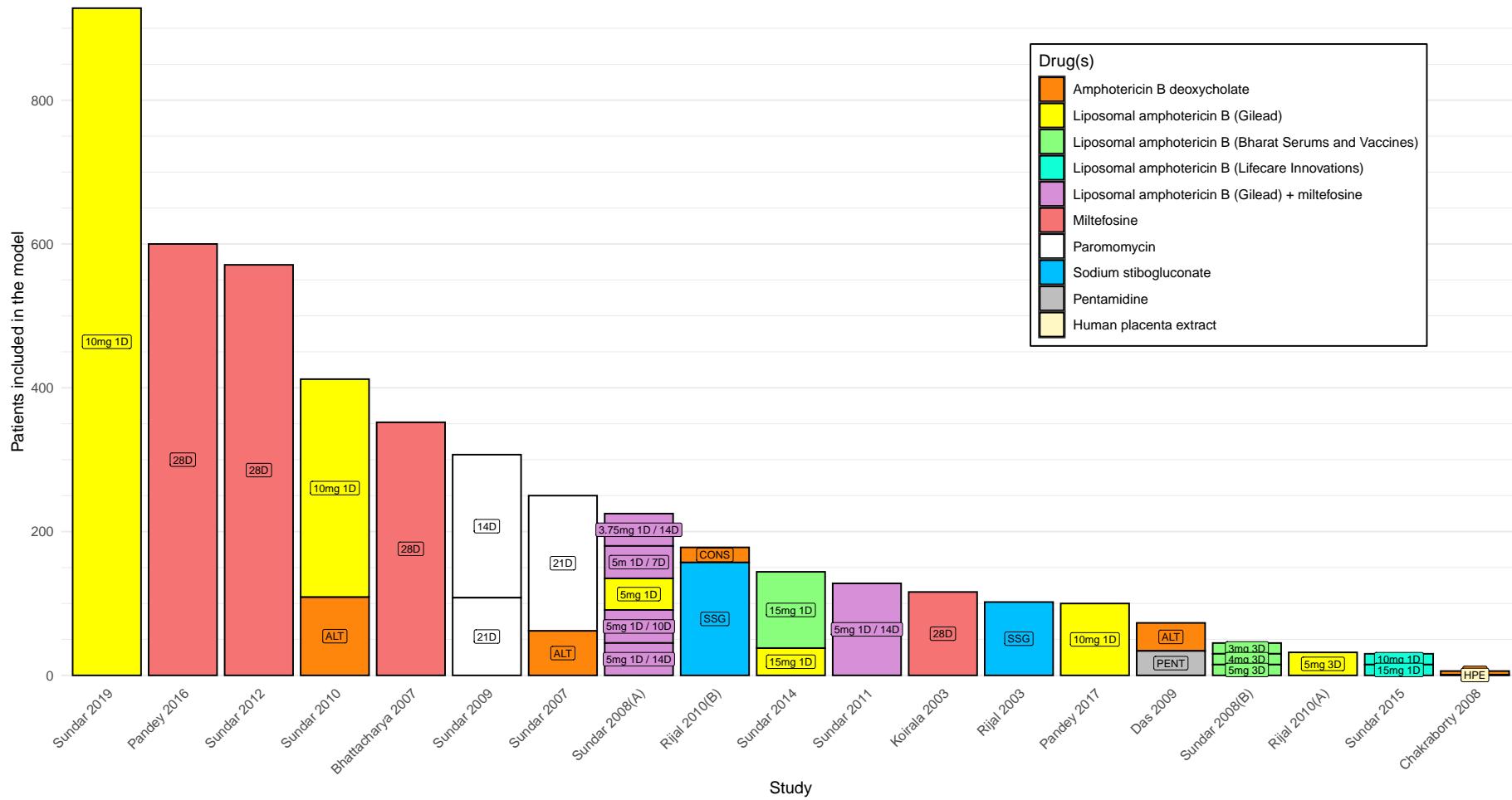


Figure 5.2: Bar chart showing the distribution of treatment regimens across contributing studies from the Indian subcontinent. Important distinguishing dosing information are provided in the overlaid labels, as space allows. Full treatment details are presented in Table 5.1. 6 patients are included from Chakraborty 2008: 5 receiving alternate day amphotericin B deoxycholate and 1 receiving human placenta extract. ABD: amphotericin B deoxycholate; ALT: alternate days; CONS: consecutive days; D: days; HPE: human placenta extract; MF: miltefosine; PENT: pentamidine; mg: milligrams/kilogram; SSG: sodium stibogluconate.

Study Design and Treatment Arms

Study designs are summarised in Table 5.1 as described in the publication or protocol. Four studies (428 [9.3%] patients) were described as Phase 2 trials[273, 274, 277, 279], four studies (1,113 [24.2%] patients) as Phase 3 trials[71, 275, 276, 278], two studies (468 [10.2%] patients) as Phase 4 trials[64, 267], and one study (32 [0.7%] patients) as a Phase 2/3 trial[271].

Ten studies (1,625 [35.3%] patients) allocated patients to more than one treatment arm[71, 265, 266, 272–276, 278, 279], of which six studies (1,411 [30.7%] patients) were randomised[71, 266, 275, 276, 278] or partially randomised[273].

Treatment arms are summarised by study in Figure 5.2. The three largest studies were all monotherapy trials that treated with either 10 mg/kg single dose liposomal amphotericin B (1 study, 928 [20.2%] patients)[94], or 28-day miltefosine (2 studies, 1,171 [25.5%] patients).

Figure 5.2 presents the different treatment regimens across all studies.

Study Eligibility Criteria

All studies recruited both male and female patients, and all but three studies restricted inclusion by age. Four studies only recruited patients aged 12 years and above (314 [6.8%] patients)[271, 273, 274, 279]. Where specified, the *lower* age limit was otherwise between 2 and 6 years. One study (100 [2.2%] patients) recruited children only (< 15 years)[269]. Otherwise, the *upper* age limit ranged between 60 and 70 years.

In the 16 studies that included baseline haemoglobin level in their eligibility criteria, the median lower threshold was 50 g/L (range 35 to 60 g/L).¹ Almost all studies excluded patients with serious illness or the presence of significant co-existing diseases. No study explicitly excluded patients based on malnutrition severity.

Twelve studies (2,021 [43.9%] patients) explicitly reported performing HIV testing on all recruited patients[64, 66, 71, 265, 267, 271, 273–276, 278, 279], and patients with confirmed VL/HIV co-infection were excluded from all but two studies (280 [6.1%] patients)[270, 272]. Patients with HIV/VL co-infection from these two studies were excluded at the IPD level, where reported.

Inclusion criteria based on prior VL treatment were reported in 12 studies (2,629 [57.2%] patients)[71, 94, 266, 270–272, 274–279], with timing since last VL treatment ranging from 10 days[274] to lifelong[94]. One study (73 [1.6%] patients) only recruited patients who had failed treatment with SSG[266].

¹with one study reporting boosting haemoglobin with blood transfusions prior to recruitment[266].

Specific wording of the inclusion and exclusion criteria, and exclusion thresholds for platelets, white blood cells, clotting, renal, and liver function, are provided in the [Supplementary Material](#).

Diagnostic Criteria

The majority of studies specified clinical criteria for inclusion, with six studies (1,984 [43.1%] patients) requiring ≥ 2 weeks' fever and splenomegaly[[94](#), [268](#), [270–272](#), [278](#)]. The remaining studies either did not specify the clinical criteria (two studies, 36 [0.8%] patients)[[265](#), [279](#)], or referred more generally to the presence of typical signs and symptoms (11 studies, 2,579 [56.1%] patients).

VL was confirmed by tissue aspirate in all but two studies, where the rK39 RDT was instead used as the primary diagnostic method (1,528 [33.2%] patients)[[94](#), [268](#)]. Three studies (276 [6.0%] patients)[[269](#), [271](#), [278](#)], screened patients with the rK39 RDT prior to confirmation with a tissue aspirate.

Relapse

Nine studies directly defined relapse in their publications or protocols (2,131 [46.3%] patients)[[64](#), [71](#), [94](#), [266](#), [267](#), [269–272](#)]. In the remaining studies, relapse events could be inferred indirectly from definitions of initial cure and treatment failure/success. Confirmation with tissue aspirates was performed in 13 studies, and reported, at least partially, in the IPD of 12 studies (see [Supplementary Material](#) for study-specific details).

In the three studies that reported relapse beyond 6 months[[94](#), [272](#), [273](#)], IPD interrogation allowed identification of the subset of patients where relapse occurred within 6 months of initial cure assessment. Active follow-up strategies were used to identify relapse events in all but one study (928 [20.2%] patients)[[94](#)].

Initial Cure

Initial cure was assessed 28–30 days after treatment initiation in 14 studies (3,740 [81.3%] patients)[[64](#), [67](#), [94](#), [265–270](#), [272](#), [276–279](#)]. In the remaining five studies, initial cure was assessed either at two different time points, depending on the treatment arm (two studies, 532 [11.6%] patients)[[71](#), [275](#)], or between days 16 and 19 (three studies, 302 [6.6%] patients)[[271](#), [273](#), [274](#)]. Clinical criteria for initial cure were described in all but two studies[[64](#), [265](#)], although often loosely defined (see [Supplementary Material](#) for study-specific details).

Tissue aspirate formed part of the initial cure assessment in all but one study (928 [20.2%] of patients without tissue aspirate)[[94](#)]. In nine studies (1,378 [30.0%]

patients)[64, 71, 267, 270, 271, 274, 275, 278, 279], patients with an initial 1+ parasite grade underwent a repeat aspirate 10 to 60 days later. If the repeat aspirate were negative, the patient would often be considered a ‘slow-responder’, allowing progression to either definite cure or relapse.

5.1.2 Patient Characteristics

Overall (marginal) distributions of categorical and continuous variables are tabulated in Tables 5.2 and 5.3, respectively. These distributions are also displayed graphically in Figure 5.4 for bar charts of categorical variables, and Figures 5.6 and 5.5 for histograms of laboratory and non-laboratory variables, respectively. Study-specific distributions are presented for age, sex, and relapse in Figure 5.3. In the Appendix, study-specific distributions of all categorical and continuous variables are presented in Figures C.1 and C.2–C.14, respectively.

Patient numbers and proportions presented in this section exclude missing data. See Section 5.1.4 for further information on missing data.

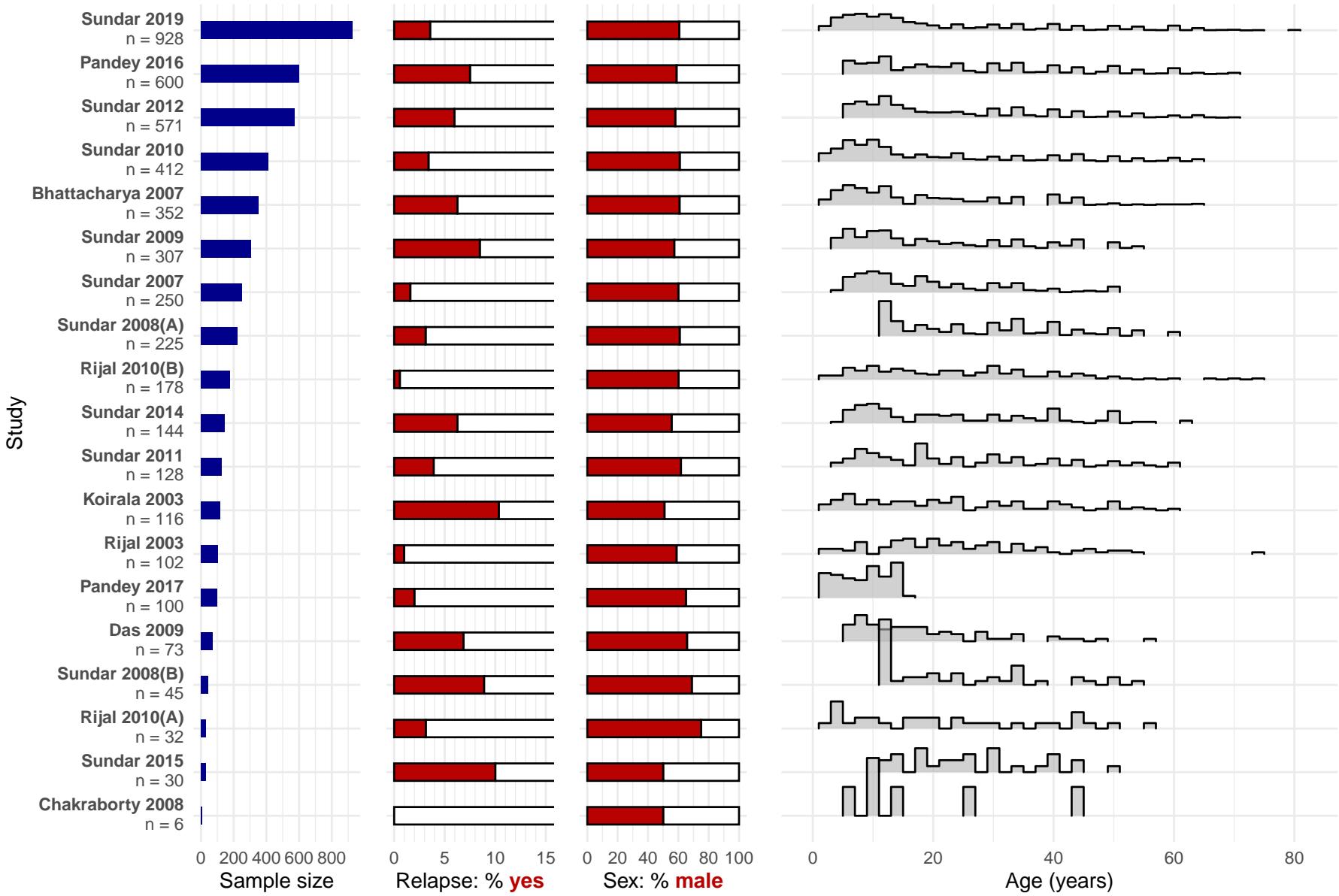


Figure 5.3: Graphical summary of the Indian subcontinent study-specific sample sizes and distributions of relapse status, sex, and age.

Categorical Variables

Across all studies, 2,745 (59.7%) of patients were male, ranging from 50.0%[265, 279] to 75.0%[271] at the study level.

Relapse within 6 months was identified in 228 (5.0%) patients, varying from 0%[265] to 10.3%[267] at the study level.

The majority of patients (52.4%) had mild/normal malnutrition according to the definition provided in Section 4.3.2. Moderate and severe malnutrition affected 28.7% and 18.9% of patients overall, although these proportions varied significantly across studies and were affected by considerable missing data (see Appendix Figure C.1).

Approximately half of all patients (2,051 [49.8%] patients) had severe anaemia at the time of recruitment, as defined by the 2024 WHO guidelines[236].

At the patient level, the most common treatment regimen was 28 days of oral miltefosine (1,639 [35.6%] patients), followed by 10 mg/kg single dose liposomal amphotericin B (1,331 [28.9%] patients). The remaining treatment regimens (1,629 [35.4%] patients), consisted of a broad range of experimental and non-experimental regimens, including amphotericin B deoxycholate, other (non-Gilead) lipid formulations of amphotericin B, paromomycin, SSG, pentamidine, and human placenta extract (Figure 5.2).

The most common baseline parasite grade on tissue aspirate was 1+ (1,201 [40.7%] patients), while the median grade was 2+ (764 [25.9%] patients at this grade; IQR: 1 to 3). The proportion of patients decreased with increasing parasite grade, with 54 (1.8%) patients with a parasite grade of 5+. Where reported, 92.0% (2,717 patients) of tissue aspirates were obtained from the spleen. The remaining aspirates were from bone marrow and accounted for almost all aspirates performed in three studies from Nepal[267, 270, 271] (Appendix Figure C.1).

Continuous Variables

The median age was 18 years, ranging from 1 to 80 years (IQR: 10 to 32 years). Overall the distribution was right-skewed, with study-specific distributions reflecting age-specific inclusion criteria. 138 (3.0%) of patients were under 5 years, and 47 (1.0%) were 65 years or over.

In adults (≥ 19 years) the median BMI was 18.2 kg/m² (IQR: 16.4 to 20.8 kg/m²) (1,313 patients). In children (≥ 5 and < 19 years), the median BMI-for-age z-score was -1.68 (IQR: -2.67 to -0.72) (1,338 patients), and in younger children (< 5 years) the median weight-for-height z-score was -1.80 (IQR: -2.97 to -0.99) (27 patients).

Variable	Overall (%) n = 4,599	Final cure (%) n = 4,371	Relapse (%) n = 228
Sex			
Female	1,854 (40.3)	1,771 (40.5)	83 (36.4)
Male	2,745 (59.7)	2,600 (59.5)	145 (63.6)
Malnutrition			
Normal/mild	1,403 (30.5)	1,333 (30.5)	70 (30.7)
Moderate	769 (16.7)	732 (16.7)	37 (16.2)
Severe	506 (11.0)	485 (11.1)	21 (9.2)
(Missing)	1,921 (41.8)	1,821 (41.7)	100 (43.9)
Anaemia			
Non-severe	2,089 (45.4)	1,973 (45.1)	116 (50.9)
Severe	2,071 (45.0)	1,985 (45.4)	86 (37.7)
(Missing)	439 (9.5)	413 (9.4)	26 (11.4)
Treatment			
Miltefosine ¹	1,639 (35.6)	1,526 (34.9)	113 (49.6)
Other	1,629 (35.4)	1,562 (35.7)	67 (29.4)
LAMB ²	1,331 (28.9)	1,283 (29.4)	48 (21.1)
Parasite grade			
1+	1,201 (26.1)	1,152 (26.4)	49 (21.5)
2+	764 (16.6)	733 (16.8)	31 (13.6)
3+	610 (13.3)	567 (13.0)	43 (18.9)
4+	323 (7.0)	301 (6.9)	22 (9.6)
5+	54 (1.2)	53 (1.2)	1 (0.4)
(Missing)	1,647 (35.8)	1,565 (35.8)	82 (36.0)
Aspirate source³			
Bone	235 (8.0)	221 (7.9)	14 (9.5)
Spleen	2,717 (92.0)	2,585 (92.1)	132 (90.4)

¹ 28 days of linear-dosed miltefosine at standard dosing.

² Single dose liposomal amphotericin B (Gilead) 10mg/kg.

³ Denominator for % in aspirate source: number of patients with documented parasite grade (overall: 2,952; final cure: 2,806; relapse: 146; no missing data).

Table 5.2: Summary of categorical candidate predictors and parasite source across contributed studies from the Indian subcontinent. Missing data are described where present. SDA: Single dose liposomal amphotericin B 10mg/kg.

The median spleen size was 4 cm (IQR: 2 to 7 cm), and ranged from 0 to 22 cm. Eight studies reported patients without splenomegaly at baseline[64, 67, 71, 94, 269, 273, 275, 276], corresponding to 218 (3.7%) patients with recorded spleen sizes.

The median duration of fever prior to patient recruitment was 30 days (IQR: 20 to 60 days). The distribution was markedly right-skewed, ranging from 1 to 730 days.

For distributions of laboratory results (haemoglobin, white blood cells, platelets, creatinine, and alanine aminotransferase), please refer to Table 5.3 and Figure 5.6.

Variable	Overall (n = 4,599)		Final cure (n = 4,371)		Relapse (n = 228)	
	Median (IQR)	Missing ¹ (%)	Median (IQR)	Missing (%)	Median (IQR)	Missing ¹ (%)
Age (years)	18 (10 – 32)	7 (0.2)	18 (10 – 33)	6 (0.1)	14 (8 – 32)	1 (0.4)
Height (cm)	150.0 (125.0 – 161.5)	1,914 (41.6)	150.0 (126.0 – 161.0)	1,815 (41.5)	149.4 (124.0 – 164.6)	99 (43.4)
Weight (kg)	37 (21 – 46)	418 (9.1)	36.8 (21 – 46)	389 (8.9)	35 (19 – 49)	29 (12.7)
BMI (kg/m^2) ²	18.22 (16.37 – 20.81)	841 (39.0)	18.17 (16.33 – 20.70)	804 (39.1)	19.77 (17.24 – 24.07)	37 (38.1)
BMI z-score ³	-1.68 (-2.67 – -0.72)	962 (41.8)	-1.67 (-2.65 – -0.71)	903 (41.5)	-1.99 (-2.81 – -1.11)	59 (47.6)
WFH z-score ⁴	-1.80 (-2.97 – -0.99)	111 (80.4)	-1.53 (-2.51 – -0.97)	108 (81.8)	-2.73 (-3.13 – -2.27)	3 (50.0)
Spleen size (cm)	4 (2 – 7)	642 (14.0)	4 (2 – 7)	595 (13.6)	3 (2 – 6)	47 (20.6)
Fever duration (days)	30 (20 – 60)	1,779 (38.7)	30 (20 – 60)	1,667 (38.1)	20 (15 – 30)	112 (49.1)
Parasite grade	2 (1 – 3)	1,647 (35.8)	2 (1 – 3)	1,565 (35.8)	2 (1 – 3)	82 (36.0)
WBC ($\times 10^9/\text{L}$)	3.4 (2.5 – 4.5)	435 (9.5)	3.4 (2.4 – 4.5)	409 (9.4)	3.4 (2.7 – 4.5)	26 (11.4)
Platelets ($\times 10^9/\text{L}$)	112 (77 – 155)	434 (9.4)	112 (77 – 155)	408 (9.3)	119.5 (83 – 160)	26 (11.4)
Haemoglobin (g/L)	79 (67 – 93)	433 (9.4)	79 (67 – 92)	407 (9.3)	82.5 (69 – 96)	26 (11.4)
ALT (IU/L)	31 (20 – 52)	449 (9.8)	31.2 (20 – 52)	421 (9.6)	30 (19 – 52)	28 (12.3)
Creatinine ($\mu\text{mol}/\text{L}$)	63.7 (51.3 – 79.6)	646 (14.0)	63.7 (51.3 – 79.6)	598 (13.7)	63.7 (51.3 – 76.0)	48 (21.1)

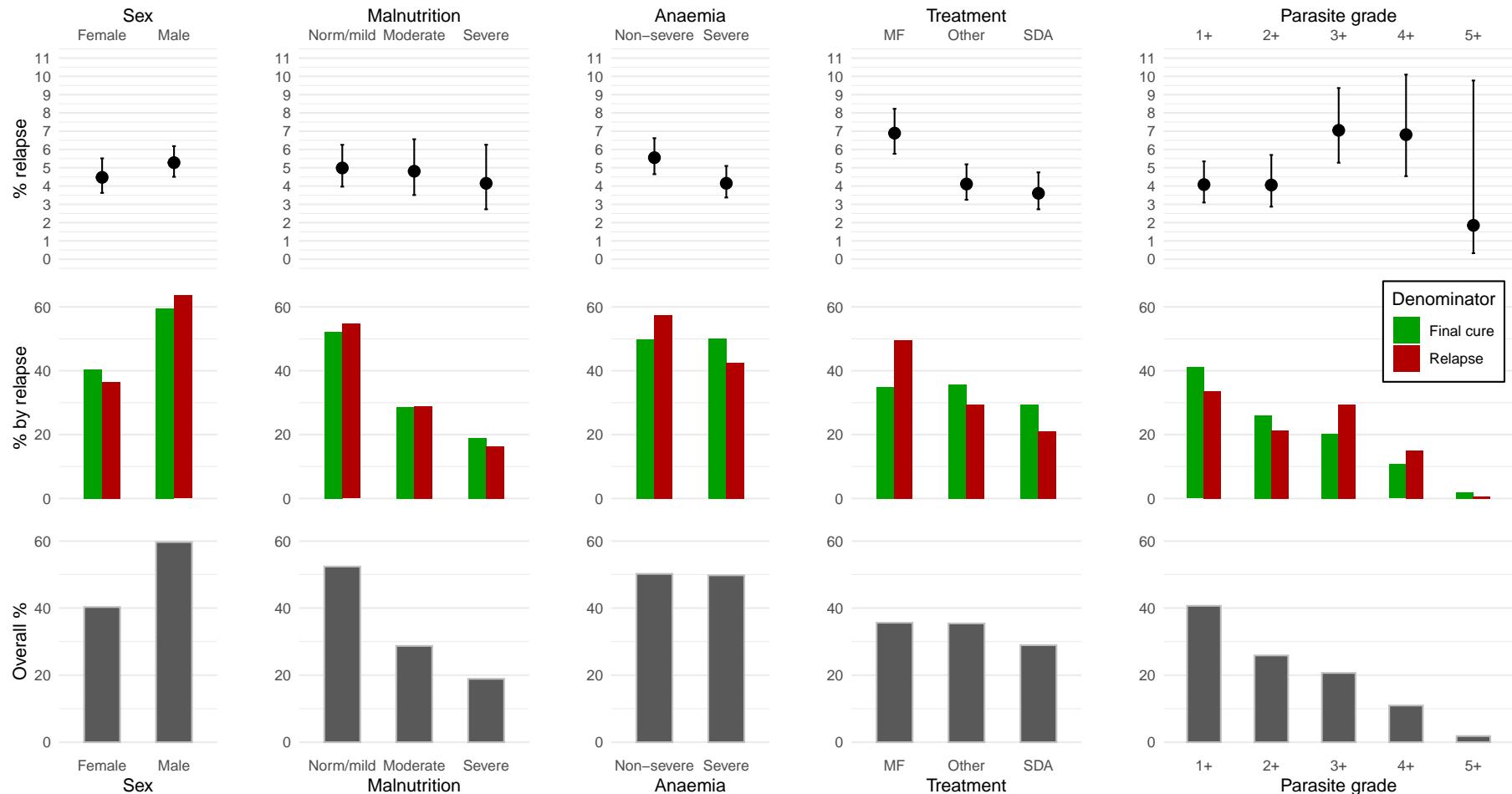
¹ Denominator for missing %: total number of patients in respective group (overall, relapse or final cure). For measures of malnutrition (BMI, BMI-for-age z-score, and weight-for-height z-score), see below.

² Denominator for missing %: number of patients aged ≥ 19 years, n = 2,154 (relapse: 97, final cure: 2,057).

³ Denominator for missing %: number of patients aged 5–18 year inclusive, n = 2,300 (relapse: 124, final cure: 2,176).

⁴ Denominator for missing %: number of patients aged < 5 years, n = 138 (relapse: 6, final cure: 132).

Table 5.3: Summary of continuous candidate predictors across contributed studies from the Indian subcontinent. Including additional variables used for the derivation of malnutrition status (height, weight, BMI, BMI-for-age z-score, weight-for-height z-score). ALT: alanine aminotransferase; BMI: body mass index; cm: centimetres; IQR: inter-quartile range, IU: international units; kg: kilograms; L: litres; m: metres; WBC: white blood cells; WFH: weight-for-height; g: grams; μmol : micromoles.



5.1.3 Univariable Associations

Unadjusted relationships between variables (including all candidate predictors, excluding missing data) and relapse risk are presented in tabular form (Tables 5.2, 5.3) and visually alongside their distributions (Figure 5.4 for categorical variables, and Figures 5.5 and 5.6 for continuous non-laboratory and laboratory variables, respectively). The relationships are presented on the log-odds (logit) scale for continuous candidate predictors in Appendix Figure C.17.

Across continuous candidate predictors, GAM smooths suggested often non-linear relationships with relapse risk, with wider uncertainty at the extremes of the predictor distributions where data were sparse. Apparent trends were most evident for age, duration of fever and spleen size. Age showed a shallow U-shaped pattern, reaching a minimum relapse risk at approximately 20 years, while duration of fever showed a marked downward trend, with longer fever durations associated with lower relapse risk. With spleen size, a notable downward trend in risk was seen for spleen sizes over 2 cm, perhaps better appreciated on the logit scale in Appendix Figure C.17. For the laboratory predictors, weak upward trending monotonic patterns were observed for white blood cells and platelets, with higher values associated with increased relapse risk.

Trends in non-candidate predictors, including weight, height, and haemoglobin, were also apparent. A marked upward trend was seen with haemoglobin and relapse risk, which was also evidence in the categorical associations, with severe anaemia associated with a lower relapse risk compared to non-severe anaemia.

Treatment with miltefosine was associated with a higher unadjusted relapse risk when compared with 10 mg/kg single dose liposomal amphotericin B or ‘Other’ treatment. Among patients who relapse, 113/228 (49.6%) were treated with miltefosine, compared to 1526/4371 (34.9%) of those who achieved final cure.

Patients with parasite grades of 3+ and 4+ were associated with higher relapse risk compared to patients with grades 1+ and 2+. Extrapolation of any trend to patients with 5+ parasite grade was limited by small numbers.

A correlation matrix showing associations between continuous variables is presented in Appendix Figure C.15, and between continuous and categorical variables in Appendix Figure C.16. These correlations are considered further in the Discussion section.

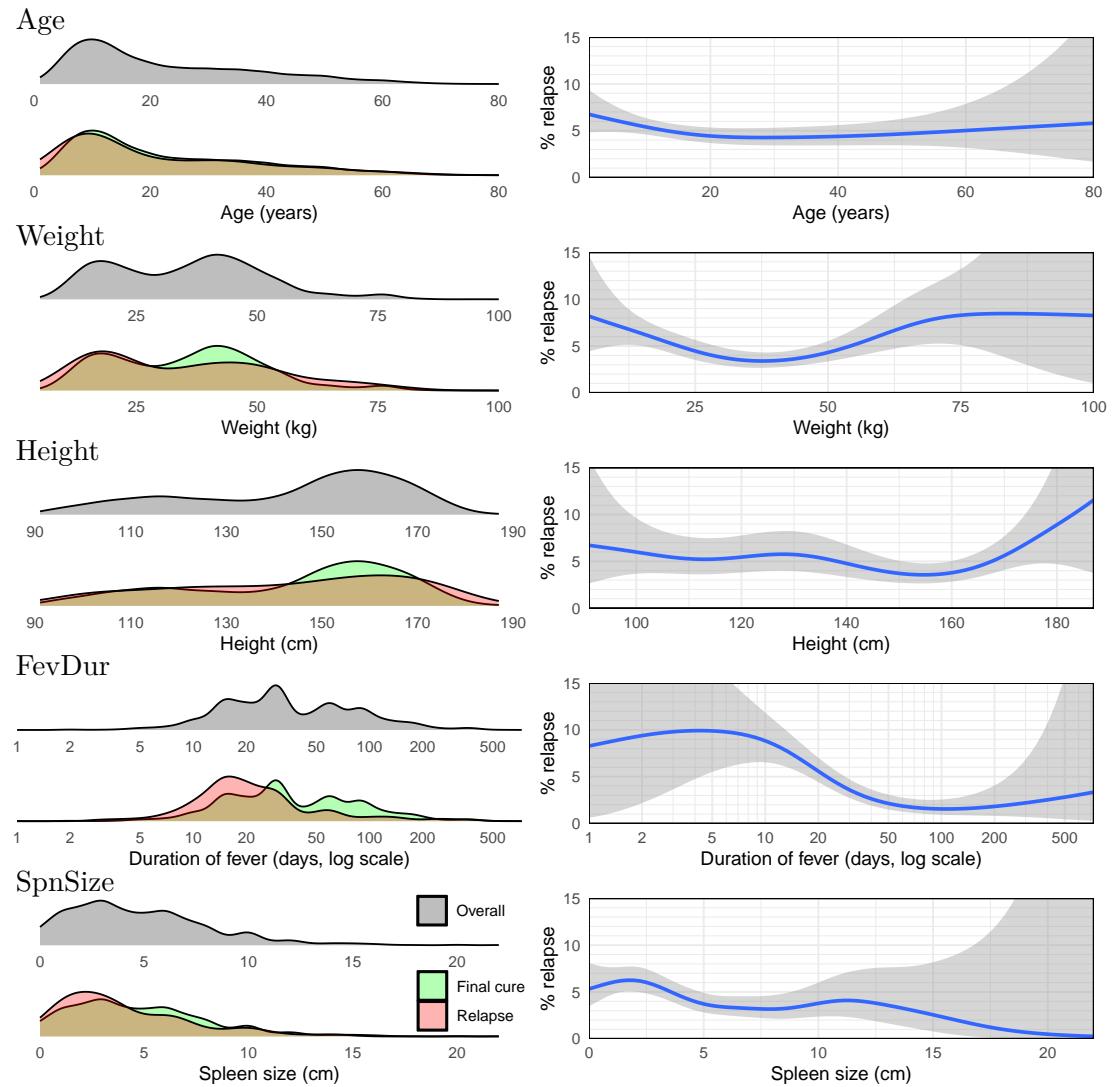


Figure 5.5: Marginal distributions and predictor-outcome relationships for continuous non-laboratory candidate predictors. FevDur: duration of fever; SpnSize: spleen size. For each candidate predictor, left upper panel shows the overall density across studies and the left lower panel shows overlapping densities normalised by relapse status. The right panel shows a univariable generalised additive model spline fit, with 95% confidence interval, of relapse.

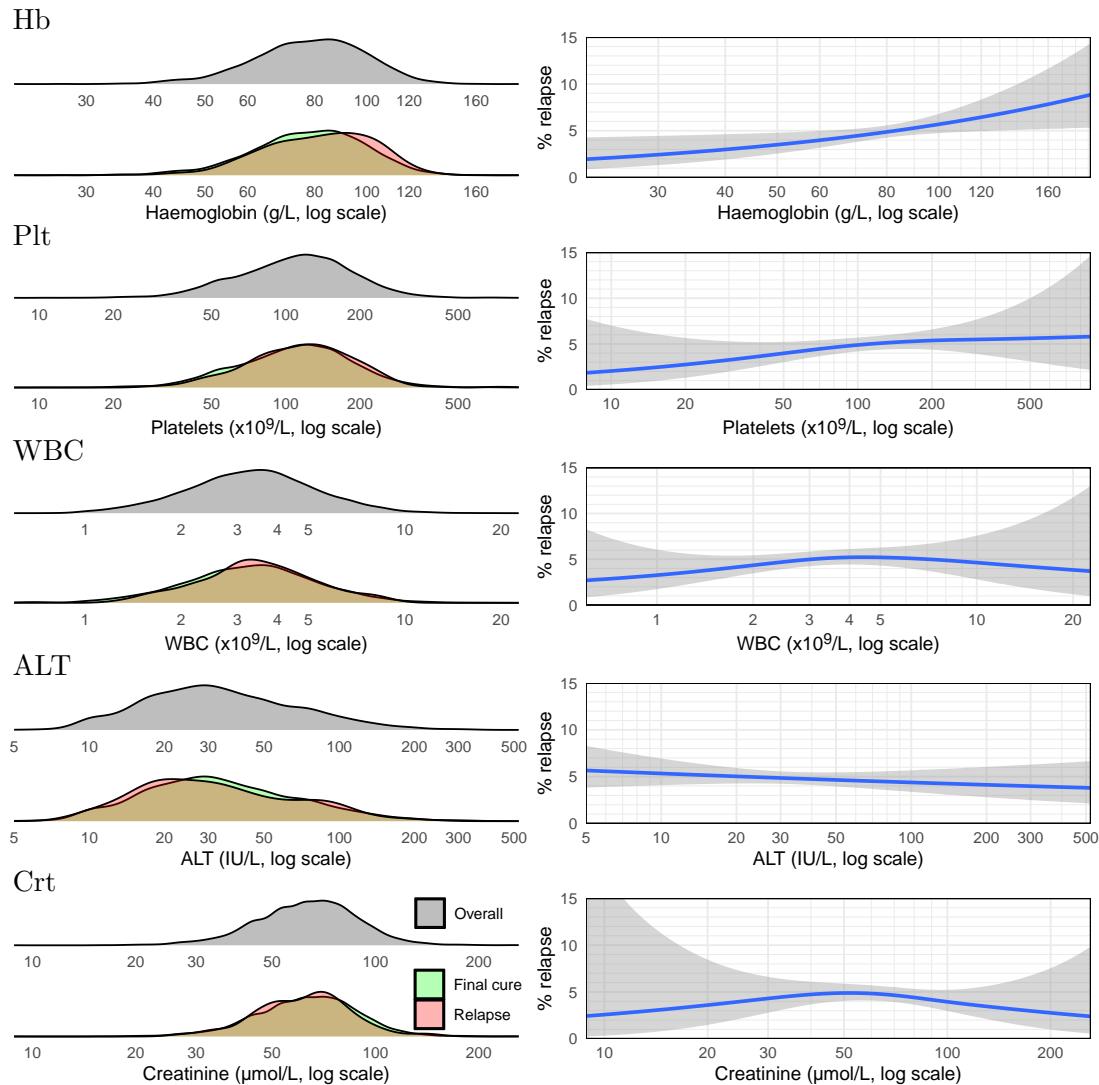


Figure 5.6: Marginal distributions and predictor-outcome relationships for continuous laboratory candidate predictors. All predictors presented on log scale. Hb: haemoglobin; Plt: platelet; WBC: white blood cells; ALT: alanine aminotransferase; Crt: creatinine. For each candidate predictor, left upper panel shows the overall density across studies and the left lower panel shows overlapping densities normalised by relapse status. The right panel shows a univariable generalised additive model spline fit, with 95% confidence interval, of relapse.

5.1.4 Missing Data

Missing data were frequent among the candidate predictors, with 3,697 (80.4%) of patients missing at least one candidate predictor data point. Just under half of all patients (2,177 [47.3%] patients) were missing exactly one data point, and 1,520 (33.1%) patients were missing two or more data points. No missing data were present in the outcome (relapse) or random-effect (study) variables.

The candidate predictor with the most missingness was malnutrition, which affected 1,921 (41.8%) patients due to a lack of height information. After malnutrition, fever duration was missing in 1,779 (38.7%) of patients, and parasite grade was missing in 1,647 (35.8%) patients. Remaining predictors had under 15% missingness. Missingness patterns, ordered by missingness in the candidate predictors, are presented in Figure 5.7 at the study level and overall.

Multiple Imputation

On review of the multiple imputation diagnostic plots (as described in Section 4.3.4 and available for review in the [Supplementary Material](#)), no clear or consistent patterns indicating violations of the missing-at-random assumption were apparent. Convergence was obtained well before the 20 iterations for the majority of imputed predictors, and the distributional assumptions of the imputation model appeared robust on inspection of the diagnostic density and scatter plots.

5.2 Model Results

Two models were fitted to the ISC IPD — one model including parasite grade as a candidate predictor, and one model excluding parasite grade.

5.2.1 Model Specification and Coefficient Estimates

Forest plots of the final model predictors are presented in Figure 5.8. Full specification of the final models, including intercept terms, p-values, and predictor transformations, are presented in Appendix Tables C.1 and C.2.

For both models, final predictors included: age (both linear and squared terms), duration of fever, treatment regimen, and presence of severe anaemia. Parasite grade was also included when considered as a candidate predictor. As can be appreciated from Figure 5.8, the adjusted odds ratios were similar across the two models. The similarity between models can also be appreciated in Figure 5.9, which shows relapse probability predictions with the model intercept recalibrated to the Sundar 2019 dataset[94].

In both models, U-shaped relationships between age and relapse risk were shown, reflecting a significant positive age squared coefficient. Inverse relationships between fever duration and relapse risk were also demonstrated. In the model including parasite grade, the relapse odds decreased by 36.0% (95% CI: 23.1–46.7%) for each doubling of fever duration, with similar findings with the model excluding parasite grade. When included, a unit increase in parasite grade (e.g. from 1+ to

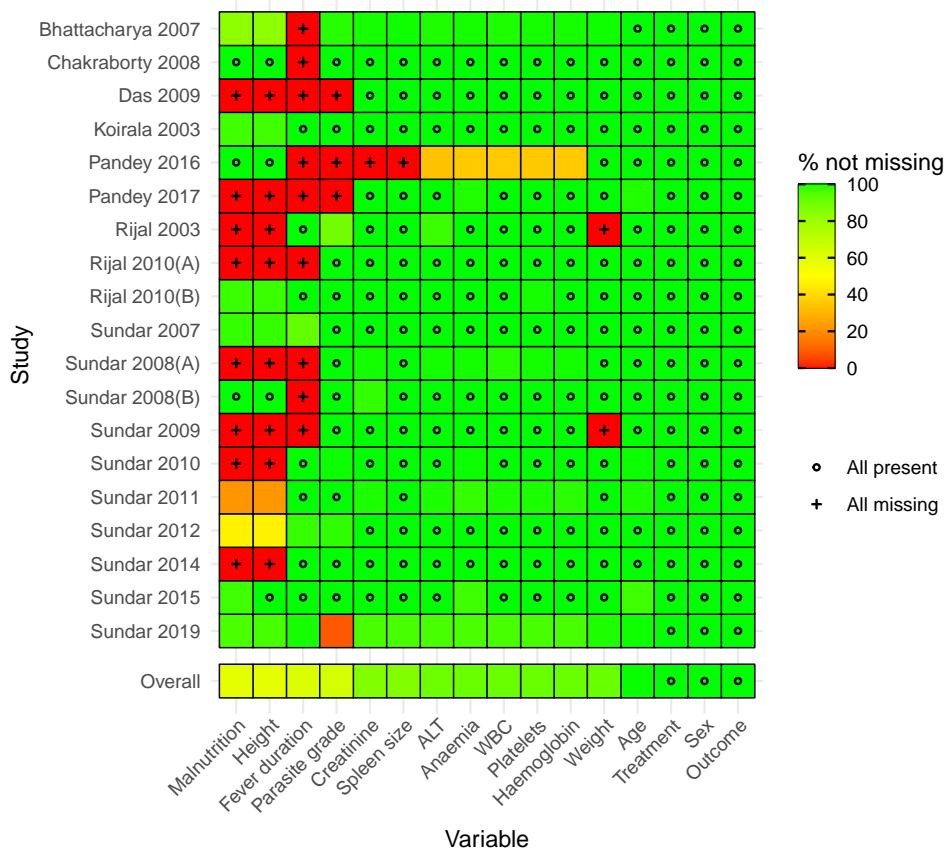


Figure 5.7: Density plot illustrating the amount of missing data overall and across contributing studies from the Indian subcontinent. Study ordered by lead author and year of publication (or protocol). Variables ordered by amount of missingness. ALT: alanine aminotransferase; WBC: white blood cells.

2+) was associated with a 30.6% increase in relapse odds (95% CI: 12.1–52.2%). Relapse odds in the 10 mg/kg single dose liposomal amphotericin B, or ‘Other’ treatment groups were significantly lower (approximately half) than the odds in the standard dose miltefosine regimen, although with marked uncertainty. The presence of severe anaemia was also found to be associated with decreased odds of relapse by approximately one third (32.0%, 95% CI: 7.0–50.3%) in the model including parasite grade. Appendix Tables C.1 and C.2 provide the complete predictor coefficients and standard errors for both models.

Pooled estimates of intraclass correlation coefficients (ICC)² were 3.5% for the model including parasite grade, and 4.4% for the model excluding parasite grade.

²ICC quantifies the proportion of total variance in the latent propensity for the outcome attributable to between-study differences.

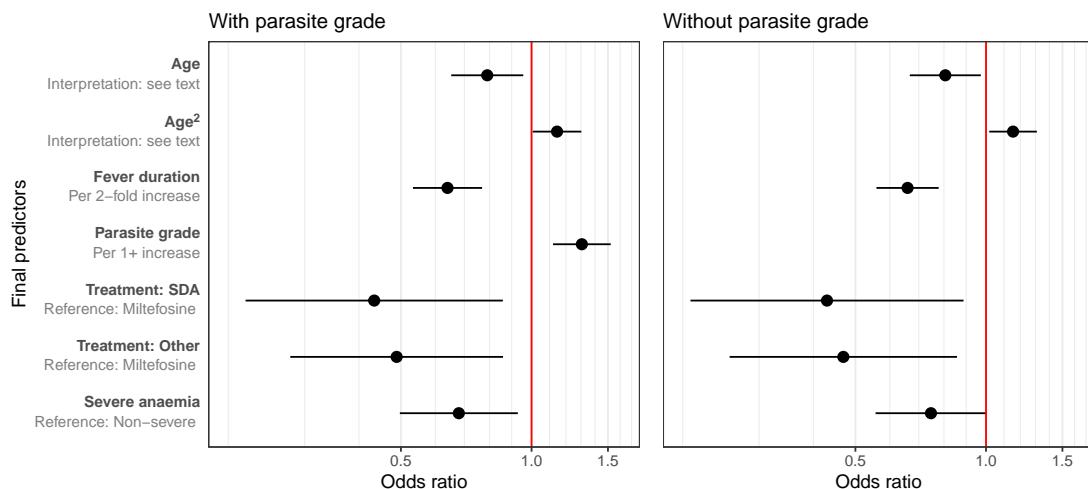


Figure 5.8: Forest plots of adjusted odds ratios with 95% confidence intervals for final model predictors. For age, the odds ratio represents a combination of linear and squared effects, with age centred by the mean and scaled by the standard deviation. Please refer to Figure 5.9 for the adjusted relapse probabilities after model intercept recalibration to the observed relapse rate in the Sundar 2019 dataset[94]. SDA: single dose liposomal amphotericin B 10 mg/kg.

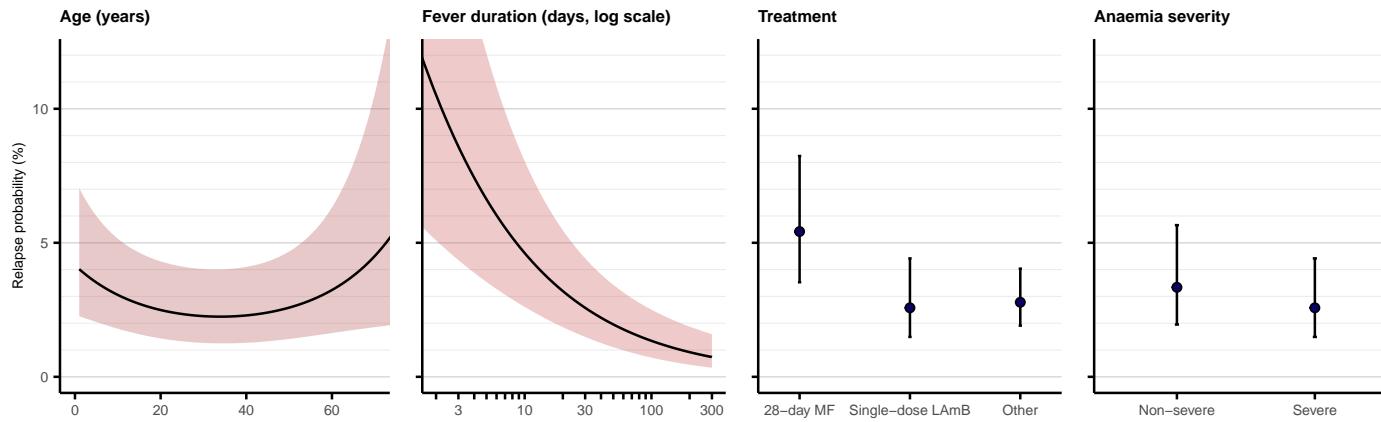
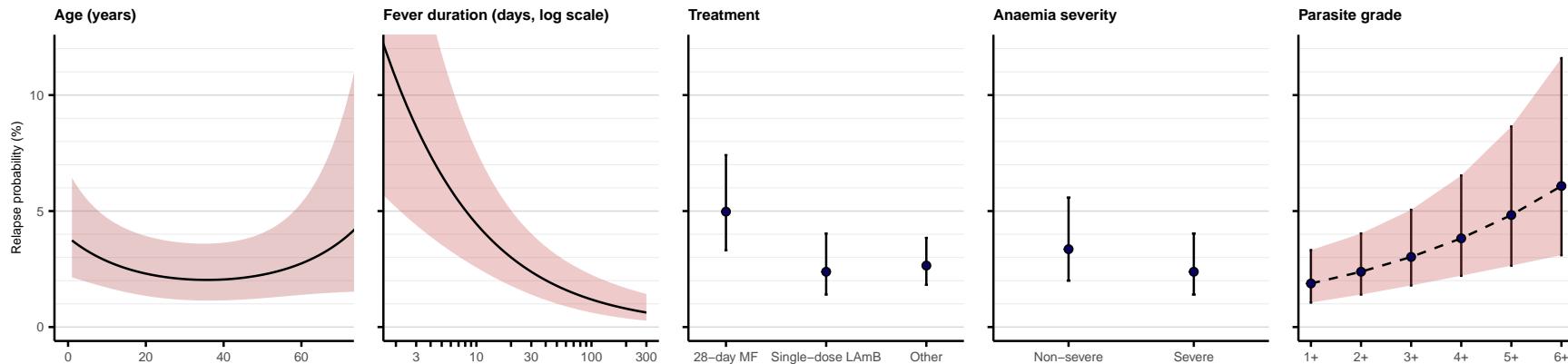
ISC model: without parasite grade**ISC model: with parasite grade**

Figure 5.9: Adjusted associations between final predictors and predicted relapse probability, as estimated from the final ISC prognostic models. Probabilities were calculated from optimism-adjusted models and following logistic recalibration (intercept-term only) to data contributed from Sundar 2019[94]. Where not varying in the plot, predictions are standardised to a representative reference participant: median age (18 years), median fever duration (30 days), treated with 10 mg/kg single dose liposomal amphotericin B, with severe anaemia, and — for the model including parasite grade — a median parasite count of 2+.

5.2.2 Model Performance and Internal Validation

Apparent and optimism-adjusted c-statistics and calibration slopes are presented in Table 5.4. When assessed in studies with >5 relapse events, model discrimination was estimated at 0.70 (95% CI: 0.66–0.74) in the model including parasite grade, and 0.69 (95% CI: 0.64–0.74) in the model excluding parasite grade. These estimates reduced to 0.68 and 0.67, respectively, after adjusting for optimism. No evidence of significant between-study heterogeneity in discrimination was identified in either model, with forest plots presented in Figure 5.10 (model including parasite grade) and Appendix Figure C.18 (excluding parasite grade).

	Estimate (95% CI)	Average optimism	Optimism-adjusted performance
Model: with PG			
C-statistic	0.70 (0.66–0.74)	0.014	0.68
Calibration slope	1.01 (0.69–1.32)	0.093	0.91
Model: without PG			
C-statistic	0.69 (0.64–0.74)	0.016	0.67
Calibration slope	1.02 (0.68–1.35)	0.098	0.92

Table 5.4: Apparent and optimism-adjusted performance measures. Abbreviations: CI: confidence interval; PG: parasite grade

Optimism in calibration slopes were estimated at 0.093 and 0.098 in the models including and excluding parasite grade, respectively. Resulting uniform shrinkage factors were 0.91 and 0.92, respectively. Multiplying together these shrinkage factors with the estimated model coefficients will result in the optimism-adjusted coefficients (Appendix Tables C.1 and C.2). No significant evidence of between-study heterogeneity in calibration slope was identified, with forest plots presented in Figure 5.11 (model including parasite grade) and Appendix Figure C.19 (model excluding parasite grade).

Calibration intercepts (calibration-in-the-large) were found to vary significantly between studies for both models, as can be appreciated from the forest plots in Figure 5.11 for the model including parasite grade, and Appendix Figure C.19 for the model excluding parasite grade (test for heterogeneity; $p = 0.01$ and $p = 0.002$ for models including and excluding parasite grade).

Visual inspection of the calibration plots comparing predicted and observed relapse probabilities showed minimal deviation from perfect calibration (Figure 5.12). Calibration plots showing predicted and observed relapse probabilities for fever

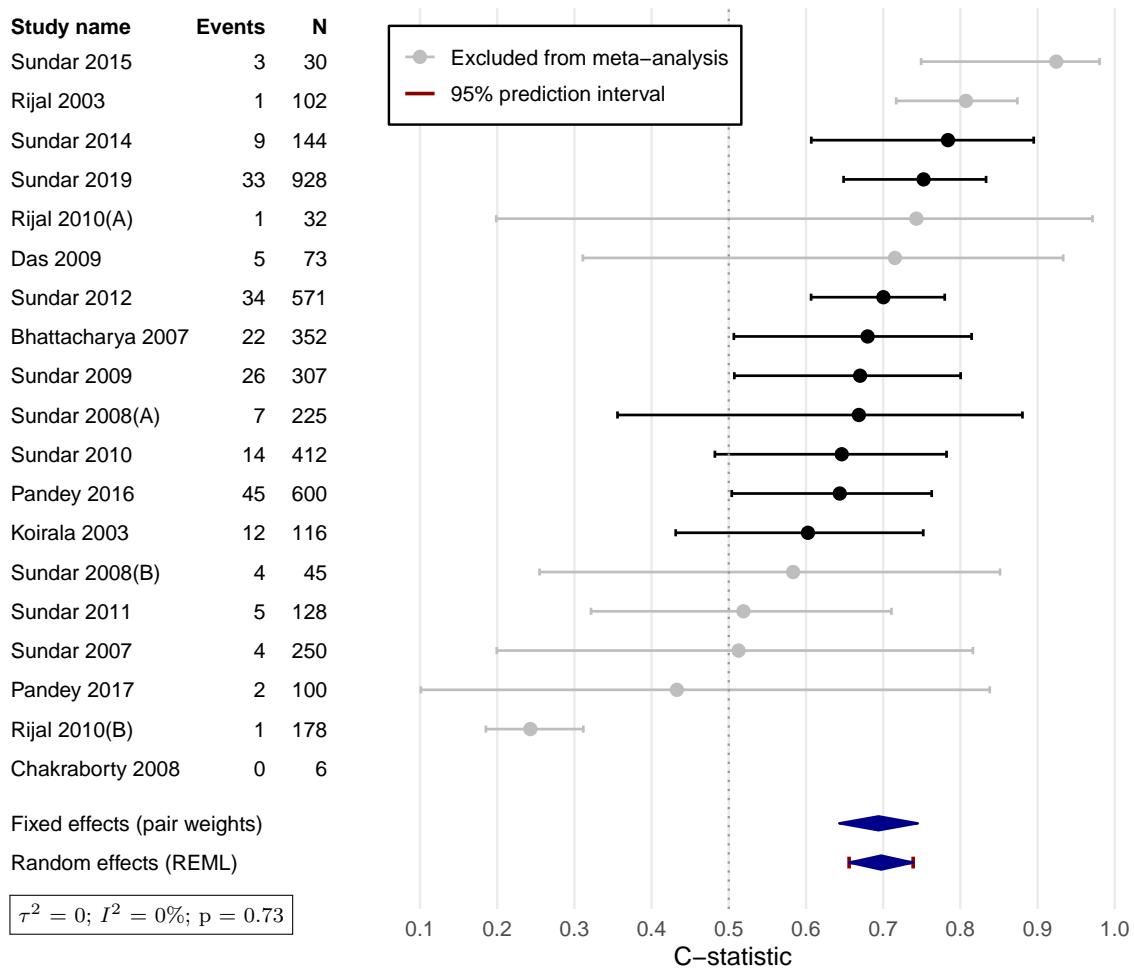


Figure 5.10: Forest plot showing individual and pooled study c-statistics, for the model **including** parasite grade. For the model excluding parasite grade refer to Appendix C.18. Pooled c-statistics are presented from both fixed-effects and random-effects meta-analysis models, after excluding studies with ≤ 5 relapse events. Blue diamonds: pooled summary estimates with 95% confidence intervals. For Chakraborty 2008, no relapse events occurred and c-statistic is therefore undefined.

duration (Figure 5.13 for model including parasite grade), and other predictors (Appendix Figures C.20 to C.24 for model including parasite grade and C.25 to C.28 excluding parasite grade) and also showed good overall agreement, with predicted relapse probabilities lying within the 95% confidence intervals of the observed probabilities.

The distribution of the final predictors selected across the 2×500 bootstrap models are presented in Appendix Table C.3. Reassuringly, the most frequently selected predictors correspond to the final predictors selected in both final models. Of note, fever duration was selected across all bootstraps for both models. When considered as a candidate predictor, parasite grade was selected in 472 (94.4%) of

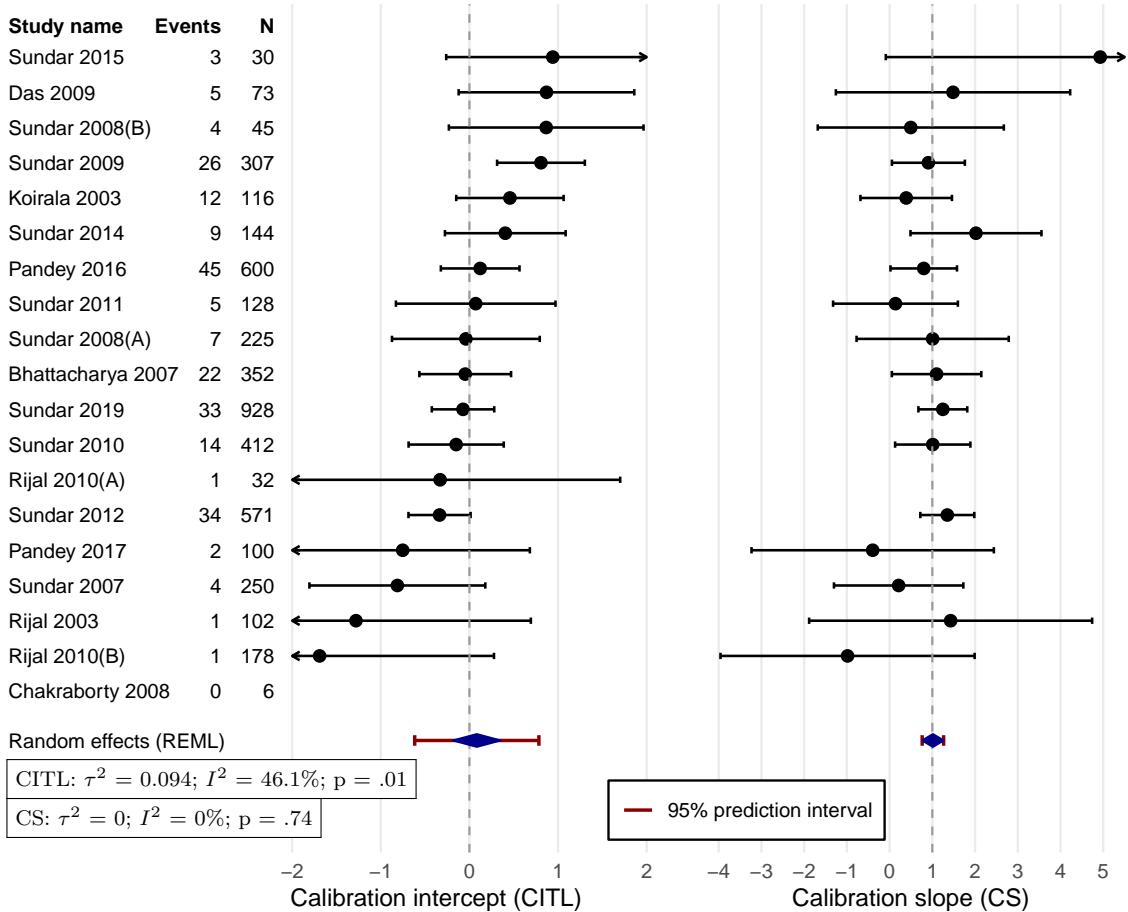


Figure 5.11: Forest plots showing individual and pooled study calibration measures for the model **including** parasite grade. Left: calibration intercept (calibration-in-the-large, CITL); Right: calibration slope (CS). Blue diamonds: summary estimates with 95% confidence intervals. Calibration measures not presented for Chakraborty 2008 due to no relapse events. Calibration slope not presented for Rijal 2010(A) due to only one relapse event and few total participants leading to failure of model convergence.

bootstrap models. A degree of instability was apparent, with predictors such as spleen size, sex, alanine aminotransferase, and the cubic age term being selected in 100–250 bootstraps.

5.3 Summary

In summary, two prognostic models predicting relapse were fitted to the ISC IPD — one including parasite grade as a candidate predictor and one excluding it. Both final models retained the same core predictors: age (modelled using linear and squared terms, yielding a U-shaped association), duration of fever (inverse association), treatment regimen (lower relapse risk with LAMB and ‘Other’ regimens

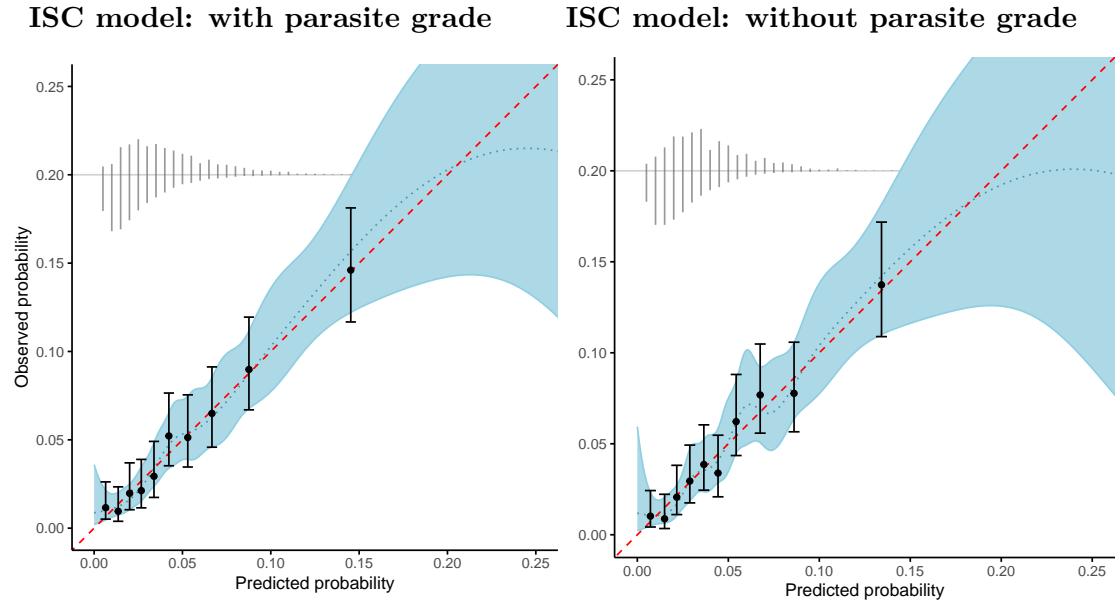


Figure 5.12: Calibration plots showing observed versus predicted probabilities for deciles of predicted probability. Red dashed line represents perfect calibration. Observed probabilities are presented with 95% confidence intervals (black error bars). A generalised additive model is fitted to show the smoothed mean observed probability (blue dotted line) with 95% confidence intervals (blue ribbon). Histograms, normalised by outcome, are overlaid to illustrate the distribution of relapses and cures across the expected probabilities.

compared with miltefosine), and severe anaemia (lower relapse risk with increased severity). When parasite grade was included, it was retained and associated with a higher risk of relapse.

Discriminative performance was modest. The pooled apparent c-statistics were 0.70 (95% CI 0.66–0.74) for the model including parasite grade and 0.69 (95% CI 0.64–0.74) for the model excluding it, reducing after optimism correction to 0.68 and 0.67, respectively. Overall agreement between observed and predicted relapse risk was good across predictors, although significant between-study variation in calibration intercepts (calibration-in-the-large) was observed for both models.

The following section presents an analogous summary of the East Africa models using the same structure, enabling direct comparison with the ISC model results.

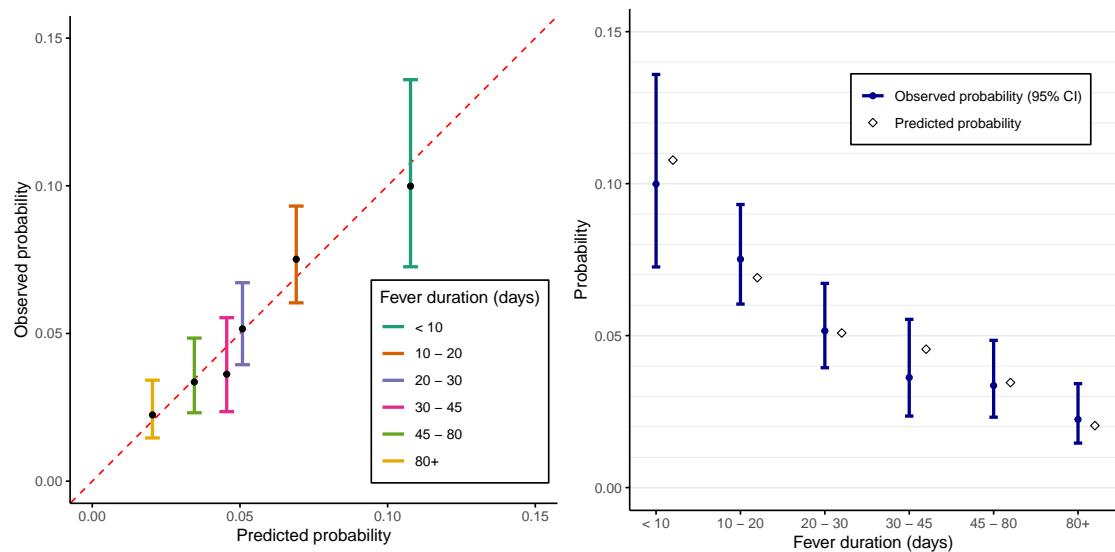


Figure 5.13: Calibration plots for different fever durations (model including parasite grade).

6

Results: East Africa Models

Contents

6.1 Descriptive Analysis	102
6.1.1 Study Characteristics	102
6.1.2 Patient Characteristics	110
6.1.3 Univariable Associations	116
6.1.4 Missing Data	118
6.2 Model Results	119
6.2.1 Model Specification and Coefficient Estimates	119
6.2.2 Model Performance and Internal Validation	122
6.3 Summary	123

6.1 Descriptive Analysis

A total of 2,095 patients from nine East African studies were identified after applying the eligibility criteria to the IDDO VL data platform[[72](#), [121–123](#), [126](#), [288–291](#)]. At the participant selection stage, 370 participants (13.5%, 370/2749) were excluded due to either having a positive HIV test, or lacking a HIV test but recruited from a site outside of Sudan (considered high-risk). 323 (11.7%, 323/2749) participants were excluded due to not achieving initial cure. A flow diagram is presented in Figure [6.1](#).

6.1.1 Study Characteristics

Important study characteristics are presented in Table [6.1](#).

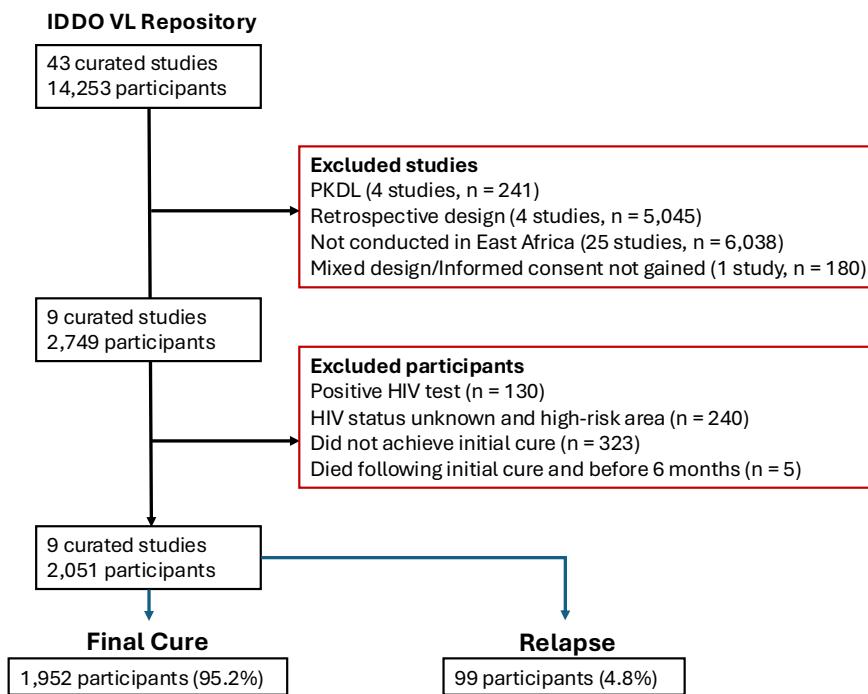


Figure 6.1: Flow diagram showing the studies and participants excluded from East Africa model development, following application of the eligibility criteria. HIV: human immunodeficiency virus; IDDO: Infectious Diseases Data Observatory; EA: East Africa; PKDL: post kala-azar dermal leishmaniasis; VL: visceral leishmaniasis.

A median of 138 patients per contributed study were included in model development, ranging from 29 to 638 patients (IQR: 92 to 289 patients). Five studies recruited from sites across multiple countries (1,186 [57.8%] patients)[[72](#), [121](#), [123](#), [126](#), [291](#)], and the remaining four studies recruited from site(s) in single countries, including Sudan (two studies, 505 [24.6%] patients)[[122](#), [290](#)] and Ethiopia (two studies, 360 [17.6%] patients)[[288](#), [289](#)]. A full breakdown of patients by country is included in Section [6.1.2](#) below.

The majority of studies were sponsored by DND*i* and conducted across the Leishmaniasis East Africa Platform (LEAP) (six studies, 1,226 [59.8%] patients, conducted between 2004 and 2016). These included LEAP 0104 — a large pivotal study comparing short course SSG & paromomycin to SSG or paromomycin monotherapy (three contributed studies, 967 [47.1%] patients)[[72](#), [122](#), [123](#)]. LEAP 0106 was a smaller study with the aim of identifying the minimum safe and effective dose of liposomal amphotericin B (LAMB) (one study, 92 [4.5%] patients)[[121](#)]. LEAP 0208 compared miltefosine with two different combination therapies (LAMB + miltefosine, LAMB + SSG) (138 [6.7%] patients)[[291](#)], and LEAP 0714 was a smaller allometric dosing study of miltefosine in paediatric patients (29 [1.4%] patients)[[126](#)].

The remaining studies were led by MSF and conducted between 1998 and 2005, including two studies comparing generic and branded SSG (577 [28.1%] patients)[[288](#), [290](#)], and a further study comparing SSG and miltefosine (248 [12.1%] patients)[[289](#)].

Table 6.1: Key characteristics of included studies from East Africa, ordered by lead author and year of publication/protocol. -: not reported; BD: twice daily (bis die); D: day(s); IM: intramuscular; IV: intravenous; LAMB: liposomal amphotericin B (Gilead formulation); MF: miltefosine; mg/kg: milligrams per kilogram; OD: once daily (omni die); PO: per os (oral); PM: paromomycin; SSG: sodium stibogluconate;

Study ¹	Title	Journal	Sponsor/ funding	Location(s)	Study design	Study arm(s) ⁴	Age (years)	Study period	n (model)	Relapses (%)
Hailu 2010[123]	LEAP 0104. Geographical variation in the response of visceral leishmaniasis to paromomycin in East Africa: a multicentre, open-label, randomized trial	PLoS Negl Trop Dis	DND <i>i</i> /MSF	5 centres: Ethiopia (Gondar University Hospital, Arba Minch Hospital); Sudan (MSF Treatment Centre, Um el Khjer; Ministry of Health Hospital, Kassab); Kenya (CCR, KEMRI, Nairobi)	Multicentre, open-label, randomised trial	(1) SSG, 20 mg/kg, IV or IM, OD, 30 D; (2) PM, 15 mg/kg, IM, OD, 21 D; (3) Combination of PM and SSG, same dose, frequency and route, 17 D	4–60	2004–2008	289	21 (7.3)
Khalil 2014[121]	LEAP 0106. Safety and efficacy of single dose versus multiple doses of AmBisome for treatment of visceral leishmaniasis in eastern Africa: a randomised trial	PLoS Negl Trop Dis	DND <i>i</i> (multiple funders)	3 centres: Ethiopia (Gondar University Hospital, Arba Minch Hospital); Sudan (Ministry of Health Hospital, Kassab)	Multicentre, open-label, non-inferiority, randomised trial with adaptive design	LAMB, IV, either single dose at (1) 7.5 mg/kg; (2) 10 mg/kg; (3) 12.5 mg/kg; (4) 15 mg/kg, or multiple dose at (5) 3 mg/kg, OD, D1–5, 14, 21 (total 21 mg/kg)	≥ 4	2009–2011	92	11 (12.0)
Mbui 2019[126]	LEAP 0714. Pharmacokinetics, Safety, and Efficacy of an Allometric Miltefosine Regimen for the Treatment of Visceral Leishmaniasis in Eastern African Children: An Open-label, Phase 2 Clinical Trial	Clin Infect Dis	DND <i>i</i> (multiple funders)	2 clinical sites: Kacheliba, West Pokot County, Kenya; Amudat, Karamoja sub-region, Uganda	Open-label, phase 2, clinical trial	(1) MF, allometric dosing according to sex, height, and weight, BD, PO, 28 D ²	4–12	2015–2016	29	2 (6.9)
Musa 2010[122]	LEAP 0104. Paromomycin for the treatment of visceral leishmaniasis in Sudan: a randomized, open-label, dose-finding study	PLoS Negl Trop Dis	DND <i>i</i> (multiple funders)	Ministry of Health Hospital, Kassab, Sudan	Open-label, phase 2, randomised, dose-finding study	(1) PM, 15 mg/kg, IM, OD, 28 D; (2) PM, 20 mg/kg, IM, OD, 21 D	4–60	2005–2006	40	8 (20.0)
Musa 2012[72]	LEAP 0104. Sodium stibogluconate (SSG) & paromomycin combination compared to SSG for visceral leishmaniasis in East Africa: a randomised controlled trial	PLoS Negl Trop Dis	DND <i>i</i> /MSF (multiple funders)	6 centres: 5 centres described in Khalil 2014 (above) and Amudat Hospital, Uganda.	Multicentre, open-label, parallel-arm, randomised trial	(1) SSG, 20 mg/kg, IV or IM, OD, 30 D; (2) PM, 20 mg/kg, IM, OD, 21 D; (3) combination of SSG, 20 mg/kg, IV or IM, OD, 17 D and PM, 15 mg/kg, IM, OD, 17 D	4–60	2004–2010	638	30 (4.7)
Ritmeijer 2001[288]	Ethiopian visceral leishmaniasis: generic and proprietary sodium stibogluconate are equivalent; HIV co-infected patients have a poor outcome	Trans R Soc Trop Med Hyg	MSF	Temporary MSF treatment centre, Densha, Ethiopia	Open-label, pseudo-randomised controlled trial	(1) SSG (generic) ³ , 20 mg/kg, IM, 30 D; (2) SSG (Pentostam, GlaxoWellcome), 20 mg/kg, IM, 30 D	all	1998–1999	112	1 (0.9)

continued on next page

Table 6.1: continued

Study ¹	Title	Journal	Sponsor/ funding	Location(s)	Study design	Study arm(s) ⁴	Age (years)	Study period	n (model)	Relapses (%)
Ritmeijer 2006[289]	A comparison of miltefosine and sodium stibogluconate for treatment of visceral leishmaniasis in an Ethiopian population with high prevalence of HIV infection	Clin Infect Dis	MSF	2 centres in Ethiopia: Humera Hospital, Mycadra Health Center	Open-label, randomised controlled trial	(1) SSG, 20 mg/kg/day, IM, OD, 30 D (extended if HIV positive); (2) MF, 100 mg/day, PO, OD, 28 D	≥ 15	2003–2005	248	6 (2.4)
Veeken 2000[290]	A randomized comparison of branded sodium stibogluconate and generic sodium stibogluconate for the treatment of visceral leishmaniasis under field conditions in Sudan	Trop Med Int Health	MSF	2 MSF treatment centres in Gedaref State, Sudan: Um Kuraa, Kassab	Open-label, pseudo-randomised controlled trial	(1) SSG (generic) ³ , 20 mg/kg, IM, 30 D; (2) SSG (Pentostam, GlaxoWellcome), 20 mg/kg, IM, 30 D	all	1998–1999	465	4 (0.9)
Wasunna 2016[291]	LEAP 0208. Efficacy and Safety of AmBisome in Combination with Sodium Stibogluconate or Miltefosine and Miltefosine Monotherapy for African Visceral Leishmaniasis: Phase 2 Randomized Trial.	PLoS Negl Trop Dis	DND <i>i</i> (multiple funders)	3 centres: Kenya (Kimale Health Centre); Sudan (Dooka Hospital and Ministry of Health Hospital, Kassab)	Phase 2, open-label, non-comparative randomised trial (adaptive-sequential design)	(1) Combination of LAMB, 10 mg/kg, IV, single dose, D1 and SSG, 20 mg/kg, IM, OD, D2-11; (2) Combination of LAMB, 10 mg/kg, IV, single dose, D1, and MF, 2.5 mg/kg, PO, OD, D2-11; (3) MF ⁵ , 2.5 mg/kg, PO, OD, 28 D	7–60	2010–2012	138	16 (11.6)

¹ Study name is composed of the lead author and year of publication.

² Refer to the supplementary material of the publication for allometric dosing table[126].

³ SSG tested and dispensed by International Dispensary Association, The Netherlands; manufactured by Albert David Ltd, India.

⁴ For both arms in Veeken 2010, Ritmeijer 2001, and Ritmeijer 2006, if positive initial test-of-cure, treatment would continue with SSG, including if previously taking MF, until two subsequent consecutive tests of cure, performed weekly, were negative.

⁵ Actual dosing ranged from 2.0–3.33 mg/kg/day after rounding to nearest 10 mg tablets; full regimen described in publication[291]

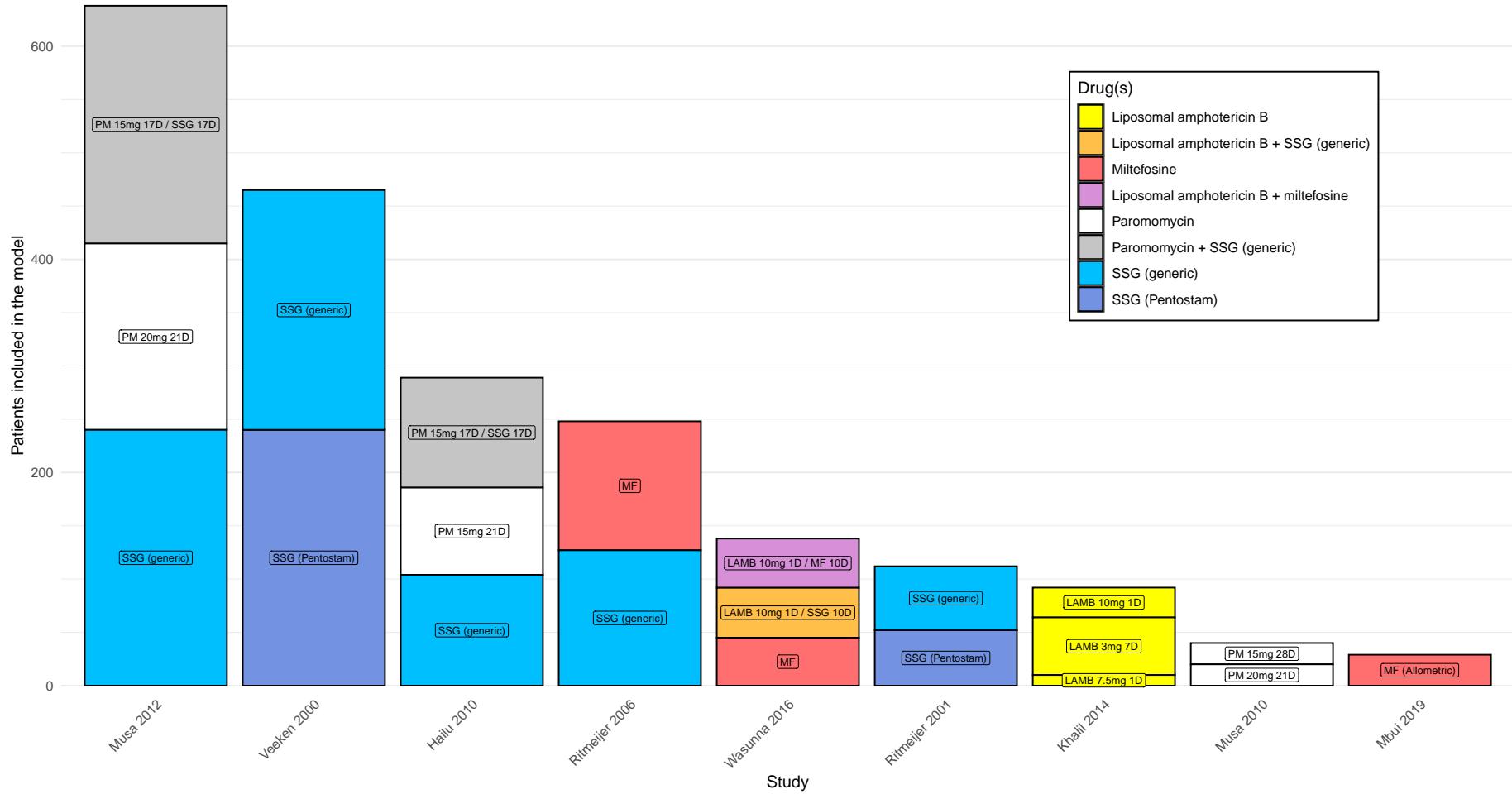


Figure 6.2: Bar chart showing the distribution of treatment regimens across contributing studies from East Africa. Drugs are colour-coded (see legend). Important distinguishing dosing information provided in the overlaid labels. Full treatment details presented in Table 6.1. D: days; LAMB: liposomal amphotericin B (Gilead); PM: paromomycin; MF: miltefosine; mg: milligrams/kilogram; SSG: sodium stibogluconate.

Study Design and Treatment Arms

All studies bar one allocated patients to more than one treatment arm (eight studies, 2,022 [98.6%] patients)[72, 121–123, 288–291]. Six studies randomly allocated patients to treatment arms (1,445 [70.5%] patients)[72, 121–123, 289, 291] and two studies were pseudo-randomised¹ (577 [28.1%] patients)[288, 290]. Three studies were described by the authors as Phase 2 trials (207 [10.1%] patients)[122, 126, 291].

Treatment arms are summarised in Figure 6.2. Treatment allocations at the patient level are described further in Section 6.1.2.

Study Eligibility Criteria

All but two studies described age limits in their inclusion criteria[72, 121–123, 126, 289, 291]. Two studies specified lower age limits only (≥ 15 years in one study; 248 [12.1%] patients, and ≥ 4 years in a further study; 92 [4.5%] patients). Three studies limited inclusion to patients between 4 and 60 years (967 [47.1%] patients)[72, 122, 123], and one study limited inclusion to patients between 7 and 60 years (138 [6.7%] patients)[291]. One study recruited children between 4 and 12 years (29 [1.4%] patients)[126].

Six studies (all DND*i*, 1,226 [59.8%] patients) reported exclusion criteria based on platelet count ($\geq 40 \times 10^9/\text{L}$) and haemoglobin ($\geq 40 \text{ g/L}$ in one study, 92 [4.4%] patients[121], and $\geq 50 \text{ g/L}$ in five studies, 1,134 [55.2%] patients[72, 122, 123, 126, 291]). Four studies described excluding patients with ‘severe’ VL (1,048 [51.1%] patients)[72, 121, 123, 126], and four studies excluded patients based on malnutrition severity (three studies excluded patients with severe protein and/or caloric malnutrition; 467 [22.8%] patients[122, 123, 291], and one study excluded children with weight-for-height or BMI-for-age z-scores < -3 ; 29 [1.4%] patients[126]). Seven studies (1,474 [71.9%] patients) excluded patients with severe concomitant illness and/or co-infection[72, 121–123, 126, 289, 291].

Four studies (299 [14.6%] patients) required a negative HIV test as a prerequisite for inclusion[121, 122, 126, 291]. The remaining studies either did not report HIV testing or HIV-related exclusion criteria (one study, 465 [22.7%] patients)[290], or permitted the inclusion of patients with VL/HIV co-infection (with variable testing strategies, 4 studies, 1,019 [49.7%] patients)[72, 123, 288, 289].

Of the eight studies that included adults (2,022 [98.6%] patients), four studies (1059/2022 [52.4%] patients) excluded women who were pregnant or lactating[72,

¹Based on whether the direct agglutination test (DAT) batch number was odd or even.

[121–123], one study excluded women of childbearing age (138/2022 [6.8%] patients)[291], one study excluded all female patients regardless of age (248/2022 [12.3%] patients)[289],² and two studies did not report excluding women based on pregnancy or lactating status(577/2051 [28.1%] patients)[288, 289].

Patients with prior VL treatment were excluded in seven studies; either within 6 months in five studies (1,088 [53.0%] patients)[72, 121–123, 126], or at any prior time point in two studies (577 [28.1%] patients)[288, 290].

Full study-specific eligibility criteria can be found in the [Supplementary Material](#).

Diagnostic Criteria

A positive tissue aspirate was required for VL confirmation in all six DND*i* studies (1,226 [59.8%] patients), with tissue type including bone marrow, spleen, or lymph node, depending on practice at the recruiting site. All studies required clinical evidence of VL, although the exact working varied across studies (presented in [Supplementary Material](#)). The three earlier MSF studies (825 [40.2%] patients)[288–290] confirmed VL with a DAT (titre \geq 1:6,400), with borderline titres (1:800–1:3,200) requiring a confirmatory tissue aspirate. One study (112 [5.5%] patients) reported some patients receiving a VL diagnosis on clinical grounds due to occasional disruption to the supply of DAT kits[288].

Relapse

Relapse was directly defined in two studies (288 [14.0%] patients)[122, 289], and could be inferred in the remaining studies through definitions of definite cure, initial cure, and treatment failure. All relapse cases were confirmed with a tissue aspirate, with aspirate results available in the contributed IPD. All studies adopted an active follow-up strategy following test-of-cure. However, loss-to-follow-up rates varied across studies, with particularly high rates seen in the earlier pragmatic MSF studies[288–290]. Four studies (1,059 [51.6%] patients)[72, 121–123] performed routine 6 month aspirates.

Initial Cure

Definitions of initial cure varied considerably across studies, and are presented in the [Supplementary Material](#). In the six DND*i* studies, test-of-cure occurred at variable time points within 31 days of treatment initiation, depending on the treatment arm (range 18–31 days)[72, 121–123, 126, 291]. In the three MSF studies, treatment

²Due to concern re: teratogenicity of miltefosine.

was extended with SSG until two consecutive tests-of-cure were negative[288–290]. Routine aspirates were performed in all studies, although two studies explicitly stated that aspirates were only performed in the presence of palpable splenomegaly or lymphadenopathy (360 [17.6%] patients)[288, 289].

In five studies (1,197 [58.4%] patients), those showing clinical improvement despite a *positive* initial test-of-cure aspirate, regardless of parasite grade, were considered to have failed the outcome of initial cure[72, 121–123, 291]. However, these patients remained in the study and could subsequently achieve definite cure at 6 months if clinical symptoms did not recur and a subsequent (e.g. 3 or 6 month) aspirate were negative. These patients, sometimes referred to as ‘slow-responders’, were included in model development despite not meeting the study definition of initial cure.

6.1.2 Patient Characteristics

Overall (marginal) distributions of categorical and continuous variables are tabulated in Tables 6.2 and 6.3, respectively. These distributions are also displayed graphically in Figure 6.4 for bar charts of categorical variables, and Figures 6.5 and 6.6 for histograms of non-laboratory and laboratory variables, respectively. Study-specific distributions are presented for age, sex, and relapse in Figure 6.3. In the Appendix, study-specific distributions of all categorical and continuous variables are presented in Figures D.1 and D.2–D.14, respectively.

Patient numbers and proportions presented in this section exclude missing data. See Section 6.1.4 for further information on missing data.

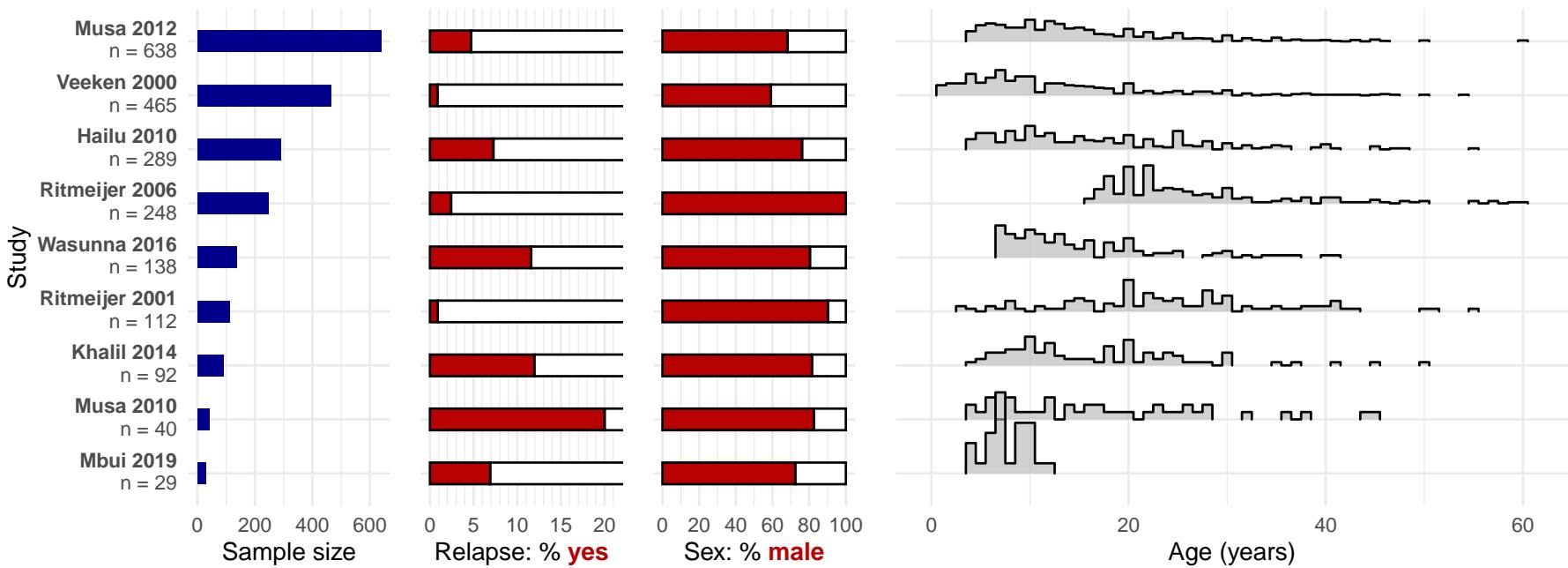


Figure 6.3: Graphical summary of East Africa study-specific sample sizes and distributions of relapse status, sex, and age.

Categorical Variables

Almost three quarters of patients (1,519 [74.1%]) were male, ranging from 59.1%[290] to 100%[289] at the study level.

Relapse within 6 months was reported in 99 (4.8%) patients overall, and varied considerably at the study level from 0.9%[288, 290] to 20%[122].

As defined in Section 4.3.2, almost a quarter of all patients were severely malnourished (509 [24.9%] patients). Moderate malnutrition affected 800 (39.1%) patients, and 735 (36.0%) had mild/normal malnutrition. Higher rates of severe malnutrition were seen in the two early MSF studies conducted in the late 1990s (33.1%)[288, 290] (presented in Appendix Figure D.1).

Severe anaemia affected 999 (48.8%) of patients overall.

A variety of treatment regimens were investigated across the studies. At the patient level, the single most common drug used was SSG (either generic or branded) — used as monotherapy in 1,048 (51.1%) patients, or as part of combination therapy with paromomycin or LAMB in 326 (15.9%) and 47 (2.2%) patients, respectively. The second most frequently used drug was paromomycin (received by 623 [30.3%] of patients overall) followed by LAMB (185 [9.1%] patients) and miltefosine (195 [9.5%] patients). The full range of treatment regimens, including dosing, are presented in Figure 6.2.

Parasite grade ranged from 1+ to 6+, with 1+ being the most common (511 [36.8%] patients who underwent tissue aspirate), followed by 2+ (237 [17.1%] patients), and decreasing to 77 (5.5%) patients with 6+. Tissue aspirate source was not reported in 702 (50.5%) of patients overall. Where reported, 393 (29.1%) were from the spleen, 163 (12.1%) were from lymph nodes, and 131 (9.7%) from bone marrow.

Approximately half of all patients were recruited from sites in Sudan (1,104 [53.8%] patients), followed by Ethiopia (695 [27.8%] patients), Kenya (227 [11.1%]), and Uganda (25 [1.2%] patients).

Continuous Variables

The median age was 14 years, ranging from 1 to 60 years (IQR: 9 to 22 years). Age distributions by study are presented in Figure 6.3, showing right-skewed distributions reflecting the study-specific age inclusion criteria. 129 (6.3%) of patients were under 5 years at the time of recruitment.

In adults the median BMI was 17.6 kg/m² (IQR: 16.3 to 18.7 kg/m²). In children (≥ 5 and < 19 years), the median BMI-for-age z-score was -2.32 (IQR: -3.10 to -1.50), and in younger children (< 5 years) the median weight-for-height z-score was -2.36 (IQR: -3.12 to -1.58) (data available for 129 patients).

Variable	Overall (%) n = 2,051	Final cure (%) n = 1,952	Relapse (%) n = 99
Sex			
Female	532 (25.9)	503 (25.8)	29 (29.3)
Male	1,519 (74.1)	1,449 (74.2)	70 (70.7)
Malnutrition			
Normal/mild	735 (35.8)	710 (36.4)	25 (25.3)
Moderate	800 (39.0)	760 (38.9)	40 (40.4)
Severe	509 (24.8)	476 (24.4)	33 (33.3)
(Missing)	7 (0.3)	6 (0.3)	1 (1.0)
Anaemia			
Non-severe	1,049 (51.1)	1,015 (52.0)	34 (34.3)
Severe	999 (48.7)	934 (47.8)	65 (65.7)
(Missing)	3 (0.1)	3 (0.2)	0 (0.0)
Parasite grade			
1+	511 (24.9)	488 (25.0)	23 (23.2)
2+	237 (11.6)	219 (11.2)	18 (18.2)
3+	192 (9.4)	181 (9.3)	11 (11.1)
4+	179 (8.7)	169 (8.7)	10 (10.1)
5+	193 (9.4)	180 (9.2)	13 (13.1)
6+	77 (3.8)	60 (3.1)	17 (17.2)
(Missing)	662 (32.3)	655 (33.6)	7 (7.1)
Aspirate source¹			
Bone	131 (9.4)	110 (8.6)	21 (22.8)
Spleen	393 (28.3)	369 (28.9)	24 (26.1)
Lymph node	163 (11.7)	151 (11.8)	12 (13.0)
(Missing)	702 (50.5)	649 (50.7)	35 (38.0)

¹ Denominator for % in aspirate source: number of patients with documented parasite grade (overall: 1,389; final cure: 1,279; relapse: 92).

Table 6.2: Summary of categorical candidate predictors and parasite source across contributed studies from East Africa. Missing data are described where present.

The median spleen size was 8 cm (IQR: 5 to 11 cm), and ranged from 0 to 30 cm. Where measured, 138 (6.9%) patients had non-palpable spleens.

The median duration of fever prior to recruitment was 40 days (IQR: 25 to 91 days), and ranged from 3 to 761 days.

Distributions of laboratory results are presented in Table 6.3 and Figure 6.5.

A correlation matrix showing associations between continuous variables is presented in Appendix Figure D.15, and between continuous and categorical variables in Appendix Figure D.16. These correlations are considered further in the Discussion section.

Variable	Overall (n = 2,051)		Final cure (n = 1,952)		Relapse (n = 99)	
	Median (IQR)	Missing ¹ (%)	Median (IQR)	Missing (%)	Median (IQR)	Missing ¹ (%)
Age (years)	14 (9 – 22)	1 (0.0)	15 (9 – 22)	1 (0.1)	12 (9 – 20)	0 (0.0)
Height (cm)	153 (126 – 168)	9 (0.4)	154 (126 – 168)	8 (0.4)	143 (127 – 164)	1 (1.0)
Weight (kg)	35 (21 – 49)	1 (0.0)	35 (21 – 49)	1 (0.1)	28 (20.6 – 44.5)	0 (0.0)
BMI (kg/m^2) ²	17.56 (16.29 – 18.69)	4 (0.5)	17.54 (16.29 – 18.71)	3 (0.4)	17.63 (16.65 – 18.22)	1 (3.2)
BMI z-score ³	-2.32 (-3.10 – -1.50)	2 (0.2)	-2.30 (-3.09 – -1.47)	2 (0.2)	-2.58 (-3.43 – -1.89)	0 (0.0)
WFH z-score ⁴	-2.36 (-3.12 – -1.58)	0 (0.0)	-2.32 (-3.01 – -1.55)	0 (0.0)	-3.42 (-3.89 – -3.24)	0 (0.0)
Spleen size (cm)	8 (5 – 11)	69 (3.4)	8 (5 – 11)	69 (3.5)	7 (4 – 10)	0 (0.0)
Fever duration (days)	40.4 (25.0 – 91.3)	349 (17.0)	40.4 (25.0 – 91.3)	320 (16.4)	30.4 (20.3 – 60.9)	29 (29.3)
Parasite grade	2 (1 – 4)	662 (32.3)	2 (1 – 4)	655 (33.6)	3 (2 – 5)	7 (7.1)
WBC ($\times 10^9/\text{L}$)	2.5 (1.8 – 3.5)	887 (43.2)	2.5 (1.8 – 3.5)	871 (44.6)	2.7 (1.8 – 3.6)	16 (16.2)
Platelets ($\times 10^9/\text{L}$)	106 (73 – 157)	890 (43.4)	105 (73 – 155)	874 (44.8)	110 (74.5 – 171)	16 (16.2)
Haemoglobin (g/L)	79 (67 – 91)	2 (0.1)	80 (67 – 92)	2 (0.1)	72 (62 – 83.5)	0 (0.0)
ALT (IU/L)	21 (14 – 31)	920 (44.9)	21 (14 – 31)	901 (46.2)	19 (14 – 30)	19 (19.2)
Creatinine ($\mu\text{mol}/\text{L}$)	61.9 (44.2 – 88.4)	826 (40.3)	61.9 (44.2 – 88.4)	815 (41.8)	61.9 (44.2 – 88.4)	11 (11.1)

¹ Denominator for missing %: total number of patients in respective group (overall, relapse or final cure). For measures of malnutrition (BMI, BMI-for-age z-score, and weight-for-height z-score), see below.

² Denominator for missing %: number of patients aged ≥ 19 years, n = 756 (relapse: 30, final cure: 726).

³ Denominator for missing %: number of patients aged 5–18 year inclusive, n = 1,165 (relapse: 65, final cure: 1,100).

⁴ Denominator for missing %: number of patients aged < 5 years, n = 129 (relapse: 4, final cure: 125).

Table 6.3: Summary of continuous candidate predictors across contributed studies from East Africa. Including additional variables used for the derivation of malnutrition status (height, weight, BMI, BMI-for-age z-score, weight-for-height z-score). ALT: alanine aminotransferase; BMI: body mass index; cm: centimetres; IQR: inter-quartile range, IU: international units; kg: kilograms; L: litres; m: metres; WBC: white blood cells; WFH: weight-for-height; g: grams; μmol : micromoles.

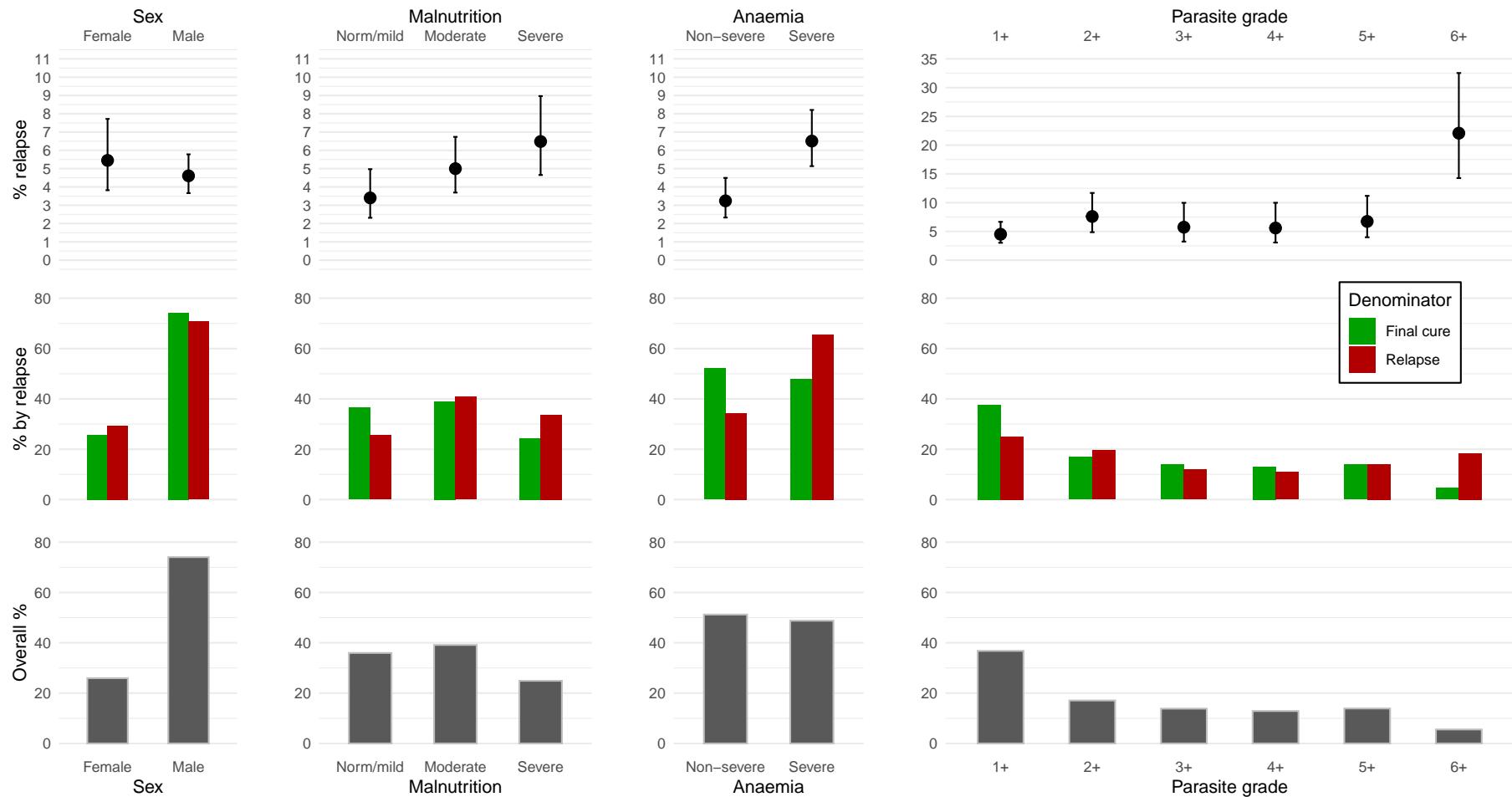


Figure 6.4: Marginal distributions and predictor-outcomes relationships for categorical candidate predictors. Excluding missing data. 95% binomial confidence intervals calculated using the Wilson method. Note: for parasite grade relapse %, y-axis is rescaled to accommodate increased risk in 6+ group. Norm: normal.

6.1.3 Univariable Associations

Unadjusted relationships between variables (including all candidate predictors) and relapse risk are presented both in tabular form (Tables 6.2, 6.3) and visually alongside their distributions (Figure 6.4 for categorical variables, and Figures 6.5 and 6.6 for non-laboratory and laboratory variables, respectively). Predictor-outcome relationships are presented on the log-odds (logit) scale for continuous candidate predictors in Appendix Figure D.17.

On review of the associations between the categorical predictors and relapse risk (upper facet row, Figure 6.4), clear trends were found for malnutrition, anaemia, and parasite grade. Increased relapse risk was seen with higher malnutrition severity, ranging from approximately 3.5% for mild/normal to 6.5% for severe categories. Perhaps more marked, with non-overlapping confidence intervals, was the increased relapse risk seen in the severe anaemia group compared to the non-severe anaemia group, from approximately 3% to 6.5%. For relapse cases, 65/99 (65.7%) had severe anaemia, compared to 476/1946 (24.5%) of final cure cases (excluding missing data).

With parasite grade, a substantial increase in relapse risk was seen for patients with a 6+ grade compared to < 6+ grades, increasing from approximately 5–7.5% to over 20%. Among relapse cases, 17/92 (18.5%) of patients had a 6+ grade aspirate, compared to 60/1297 (4.6%) of the final cure cases (after excluding missing data).

Clear trends were observed between all non-laboratory continuous variables and relapse risk (Figure 6.5). For age, relapse risk followed a non-linear trend, peaking at approximately 10 years and decreasing to a nadir in the early 20s. Fluctuations in risk with height and weight were also noted, mirroring the relationship between age and relapse. Less marked downward trends between relapse risk and spleen size and fever duration were observed.

Among the laboratory continuous variables (Figure 6.6), a pronounced inverse relationship between relapse risk and haemoglobin was demonstrated, concordant with the aforementioned increased risk seen in patients with severe vs. non-severe anaemia. Appreciable upwards trends between relapse risk and both WBC count and platelet count were also seen, although with uncertain significance.

A correlation matrix showing associations between continuous variables is presented in Appendix Figure D.15, and between continuous and categorical variables in Appendix Figure D.16. These correlations are considered further in the Discussion section.

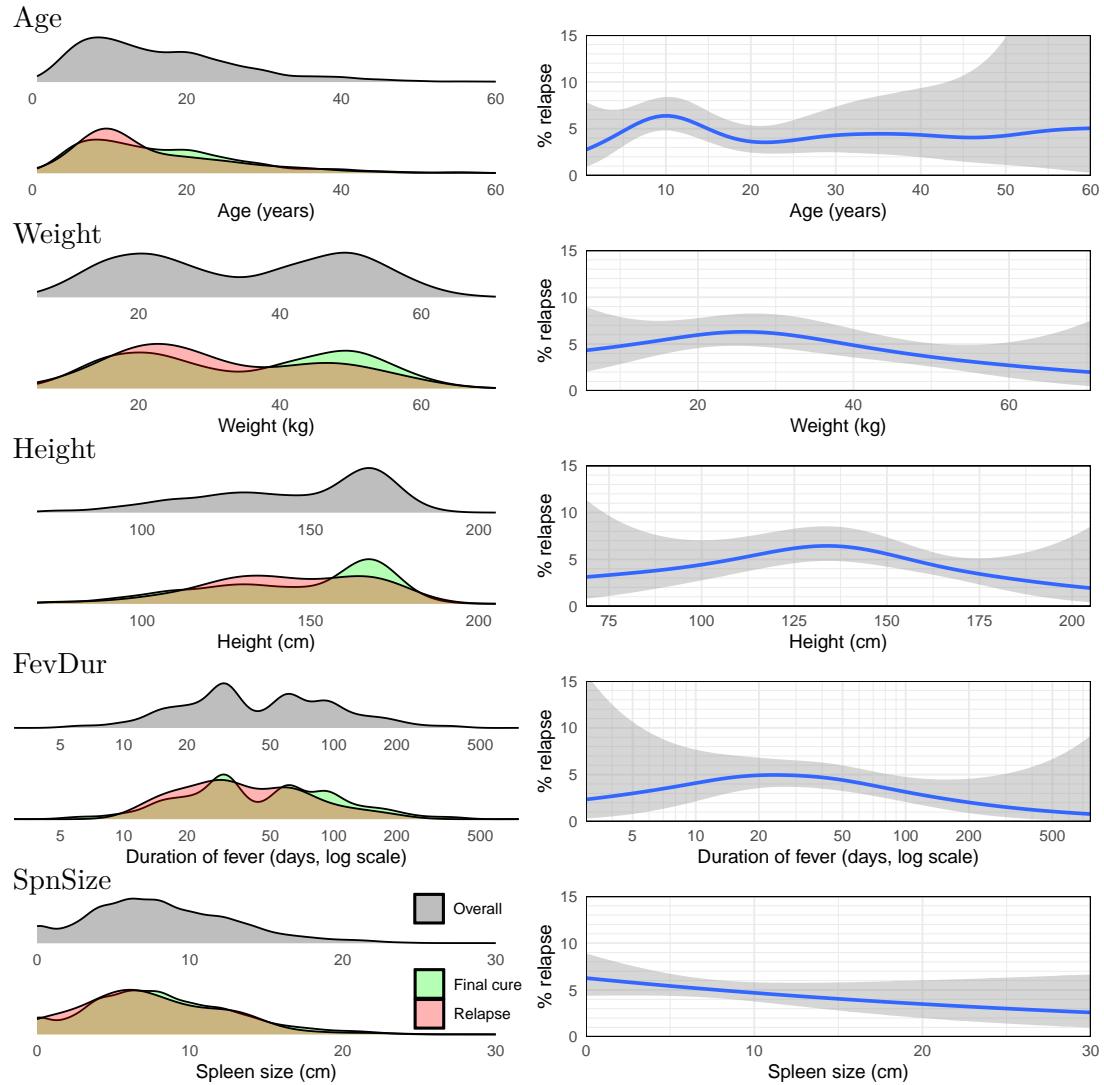


Figure 6.5: Marginal distributions and predictor–outcome relationships for continuous non-laboratory candidate predictors. FevDur: duration of fever; SpnSize: spleen size. For each candidate predictor, left upper panel shows the overall density across studies and the left lower panel shows overlapping densities normalised by relapse status. The right panel shows a univariable generalised additive model spline fit, with 95% confidence interval, of relapse.

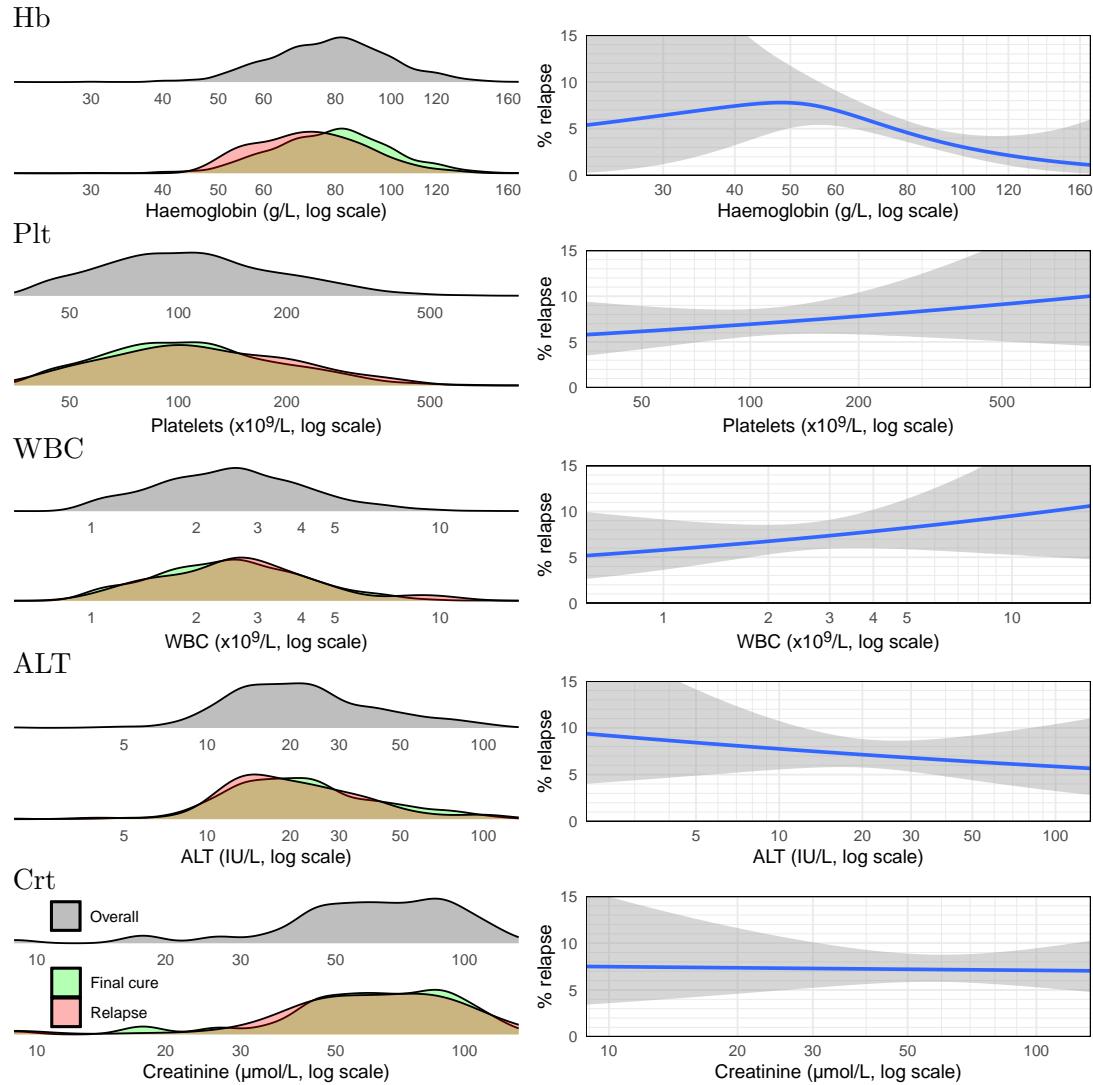


Figure 6.6: Marginal distributions and predictor–outcome relationships for continuous laboratory candidate predictors. All predictors presented on log scale. Hb: haemoglobin; Plt: platelet count; WBC: white blood cell count; ALT: alanine aminotransferase; Crt: creatinine. For each candidate predictor, left upper panel shows the overall density across studies and the left lower panel shows overlapping densities normalised by relapse status. The right panel shows a univariable generalised additive model spline fit, with 95% confidence interval, of relapse.

6.1.4 Missing Data

Missing data were common, affecting one or more candidate predictors in 1,168 (56.9%) patients. 215 (10.5%) patients were missing exactly one candidate predictor. As appreciated from Figure 6.7, the majority of missingness originated from the earlier MSF studies, where ALT, platelet count, WBC count, creatinine, and parasite grade were not routinely collected, resulting in almost one-third of all patients (634 [30.9%] patients) missing 5 candidate predictors.

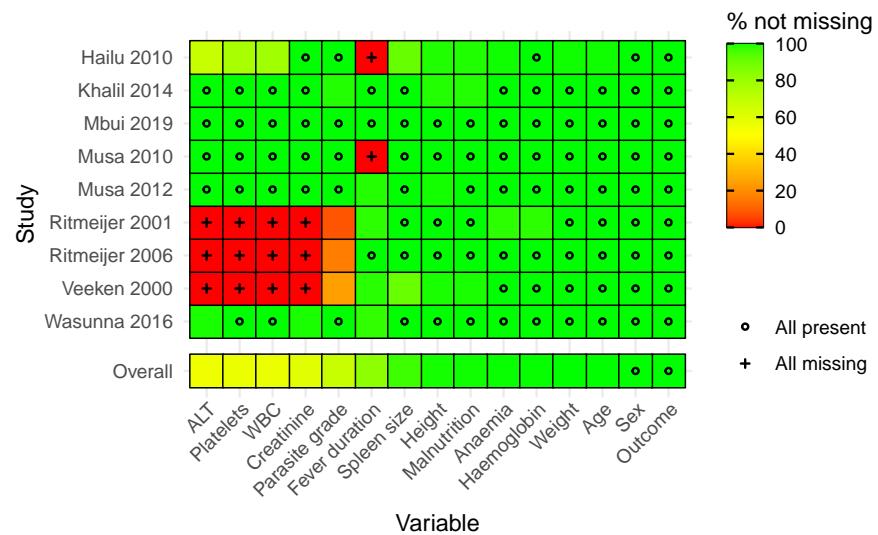


Figure 6.7: Density plot illustrating the amount of missing data overall and across contributing studies from East Africa. Study ordered by lead author and publication year. Variables ordered by amount of missingness. ALT: alanine aminotransferase; WBC: white blood cells.

Across all studies, ALT, platelet count, WBC count, and creatinine were missing in 40–45% of all patients. Parasite grade was missing for approximately one-third of the total cohort (662 [32.3%] of patients).

Multiple Imputation

Diagnostic plots for the 30 imputed datasets are available in the [Supplementary Material](#). On review of the trace plots, there was good evidence for convergence of the imputation model. Density and scatter plots revealed overall good correlation between the imputed and original variables.

6.2 Model Results

As described in the Methodology, two separate prediction models were developed from the East Africa IPD; one including parasite grade as a candidate predictor, and one excluding parasite grade.

6.2.1 Model Specification and Coefficient Estimates

Final (retained) predictors across both models are presented in Figure 6.8. Full specification of the final models, including intercept terms, p-values, and predictor transformations, are presented in Appendix Tables D.1 and D.2.

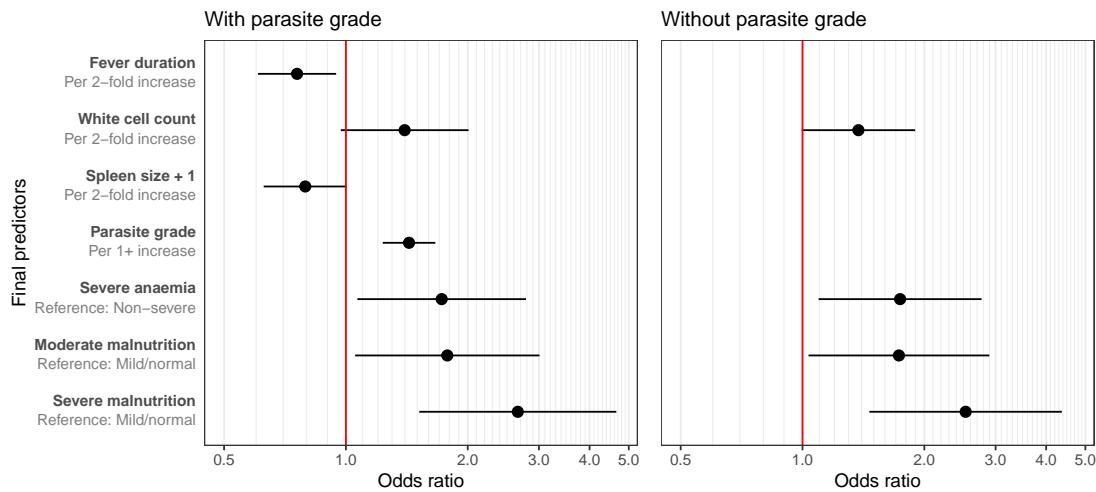


Figure 6.8: Forest plot of adjusted odds ratios with 95% confidence intervals for final model predictors. Odds ratios are displayed on a logarithmic scale.

Shared final predictors across both models included WBC count, anaemia, and malnutrition. The effect size (odds ratios) between these predictors and relapse are approximately the same across both models, and are described here for the model with parasite grade. For WBC count, the relapse odds increases by approximately 40% per doubling of the WBC count (OR: 1.40, 95% CI: 0.97–2.01). Increasing relapse odds are also seen in patients with severe vs. non-severe anaemia (OR: 1.72, 95% CI: 1.07–2.79), and in patients with moderate and severe malnutrition (OR: 2.66, 95% CI: 1.52–4.66 comparing severe vs. mild/normal malnutrition). For the similar estimates in the model without parasite grade please refer to Appendix Table D.2 and aforementioned figures.

In the model that considered parasite grade, three additional predictors were retained: parasite grade, spleen size, and fever duration. Odds of relapse were predicted to increase by over 40% per 1+ increase (OR: 1.43, 95% CI: 1.23–1.66). As observed in the unadjusted predictor-relapse relationships, inverse relationships were seen with spleen size and fever duration. Relapse risk decreased by approximately 25% per doubling of fever duration (OR: 0.75, 95% CI: 0.61–0.95) and by approximately 20% per doubling of [spleen size + 1], for example, from 2 to 5 cm, or 4 to 9 cm (OR: 0.79, 95% CI: 0.63–1.01).

As presented in Appendix Tables D.1 and D.2, considerable between-study heterogeneity was present, with pooled ICC estimates of over 20% in both models.

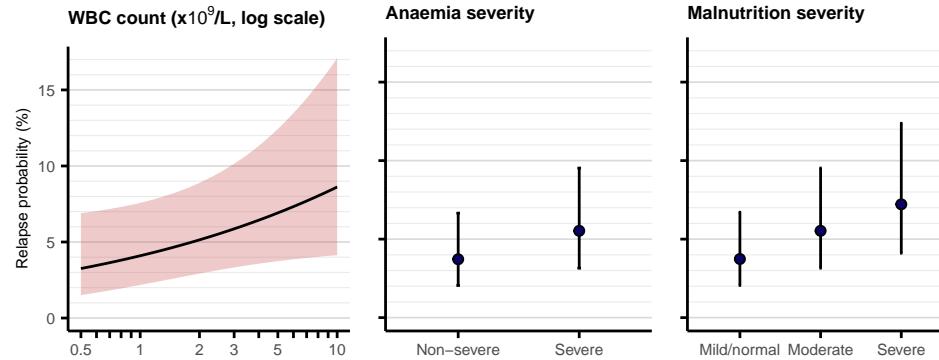
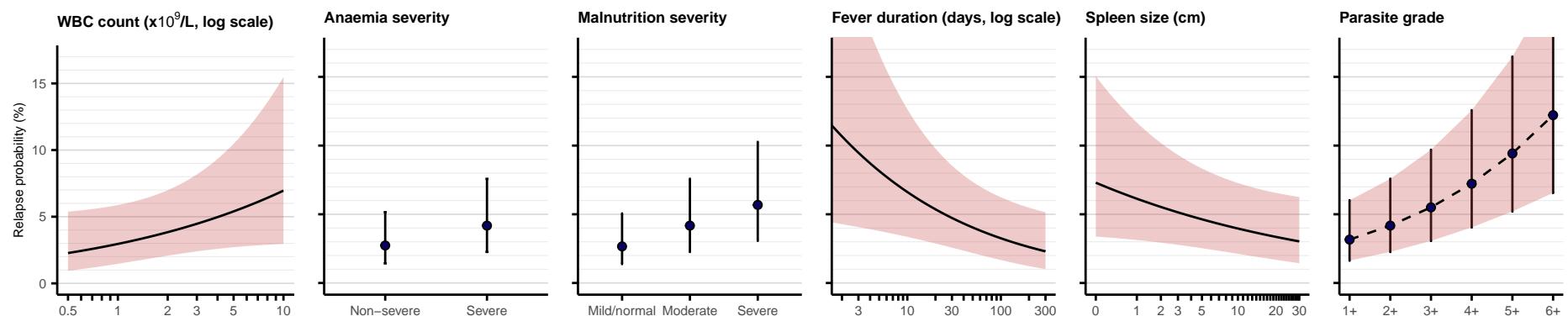
East Africa model: without parasite grade**East Africa model: with parasite grade**

Figure 6.9: Adjusted associations between final predictors and predicted relapse probability, as estimated from the final East Africa prognostic models. Probabilities were calculated from optimism-adjusted models and following logistic recalibration (intercept-term only) to data contributed from Musa 2012[72]. Where not varying in the plot, predictions are standardised to a representative reference participant: median age (13 years), median fever duration (45 days), median white blood cell count ($2.5 \times 10^9/L$), with severe anaemia, moderate malnutrition, and — for the model including parasite grade — a median parasite count of 2+.

6.2.2 Model Performance and Internal Validation

Apparent and optimism-adjusted performance measures are presented in Table 6.4. After excluding studies with ≤ 5 events, the overall c-statistics were estimated at 0.72 (95% CI: 0.63–0.82) and 0.66 (95% CI: 0.56–0.76) in the models including and excluding parasite grade, respectively. After adjusting for optimism, these estimates shrank to 0.68 and 0.61, respectively. Both models showed some evidence of between-study heterogeneity in discrimination ($I^2 = 46.6\%$, $p = 0.11$ for the model including parasite grade, and $I^2 = 53.1\%$, $p = 0.056$ for the model excluding parasite grade), as illustrated in Figure 6.10 and Appendix Figure D.18.

	Performance estimate (95% CI)	Average optimism	Optimism-adjusted performance
Model: with PG			
C-statistic	0.72 (0.63–0.82)	0.046	0.68
Calibration slope	1.03 (0.64–1.42)	0.224	0.81
Model: without PG			
C-statistic	0.66 (0.56–0.76)	0.050	0.61
Calibration slope	1.08 (0.47–1.68)	0.329	0.75

Table 6.4: Apparent and optimism-adjusted performance measures. Abbreviations: CI: confidence interval; PG: parasite grade

Optimism in calibration slope was moderate to high, estimated at 0.22 and 0.33 for the models including and excluding parasite grade, respectively. Uniform shrinkage factors were 0.81 and 0.75, respectively (Appendix Tables D.1 and D.2). There was no evidence for significant between-study heterogeneity in calibration slope.

Calibration intercepts were shown to vary significantly across the contributing studies for both models — with $I^2 = 89.0\%$ for the model including parasite grade, and $I^2 = 90.7\%$ for the model excluding parasite grade (both $p < 0.0001$ for test of heterogeneity).

Forest plots showing the heterogeneity of calibration performance measures across the contributing studies are presented in Figure 6.11 (model including parasite grade) and Appendix Figure D.19 (model excluding parasite grade).

Calibration plots demonstrated concordance between predicted and observed relapse risk across both models (see Figure 6.12). Concordance is also demonstrated when comparing predicted and observed probabilities for malnutrition severity (Figure 6.13 for the model including parasite grade and Appendix Figure D.27 for the model excluding parasite grade). Calibration plots for all final predictors are presented in the Appendix Figures D.20 to D.28.

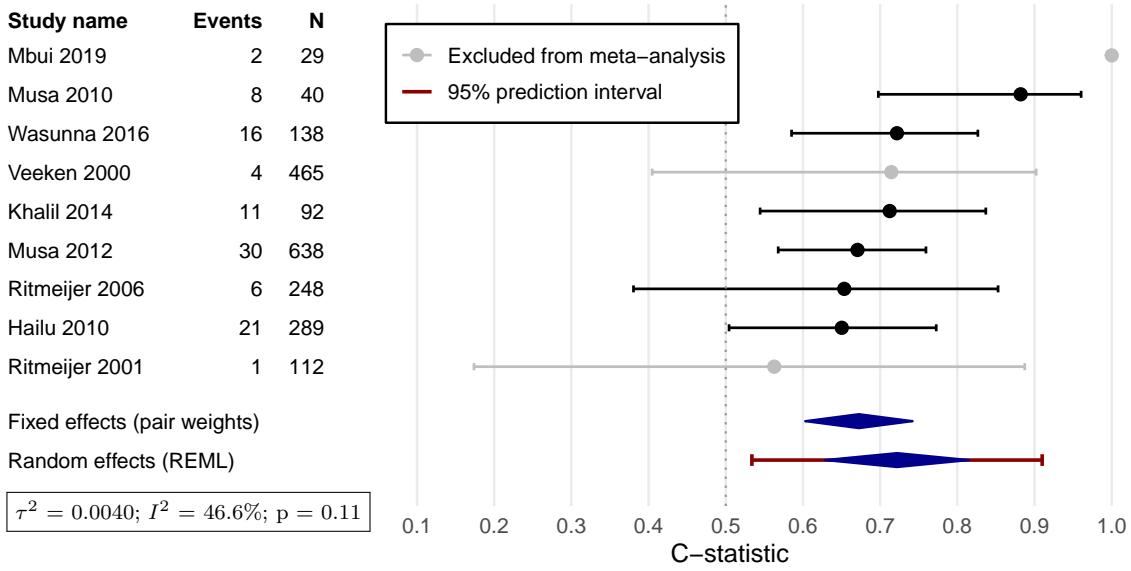


Figure 6.10: Forest plot showing individual and pooled study c-statistics, for the model **including** parasite grade. For the model excluding parasite grade refer to Appendix D.18. Pooled c-statistics are presented from both fixed-effects and random-effects meta-analysis models, after excluding studies with ≤ 5 relapse events. Study-specific confidence intervals with few events should be interpreted with caution (as discussed in the Methodology 4.3.7). Blue diamonds: pooled summary estimates with 95% confidence intervals. For Mbui 2019, all bootstrapped c-statistic estimates were 1 exactly and therefore no variance is presented.

The distribution of final predictor selection across the 2×500 bootstrap models are presented in Appendix Table D.3. When included as a candidate predictor, parasite grade consistently selected for inclusion across almost all bootstrap models (499 [99.8%] bootstrap models). In both models, the selection of malnutrition was also consistently seen (477 [95.4%] and 461 [92.2%] bootstrap models when including and excluding parasite grade, respectively). WBC count was the least stable predictor included in both models (254 [50.8%] and 188 [37.6%] bootstrap models when including and excluding parasite grade, respectively).

6.3 Summary

In summary, two prognostic models predicting relapse were fitted to the East Africa IPD — one including parasite grade as a candidate predictor and one excluding it. Both final models retained the same three core predictors: WBC count (higher relapse odds with increasing count), severe anaemia (higher relapse odds compared with non-severe anaemia), and malnutrition (higher relapse odds with increasing severity). When parasite grade was included, it was retained and strongly associated

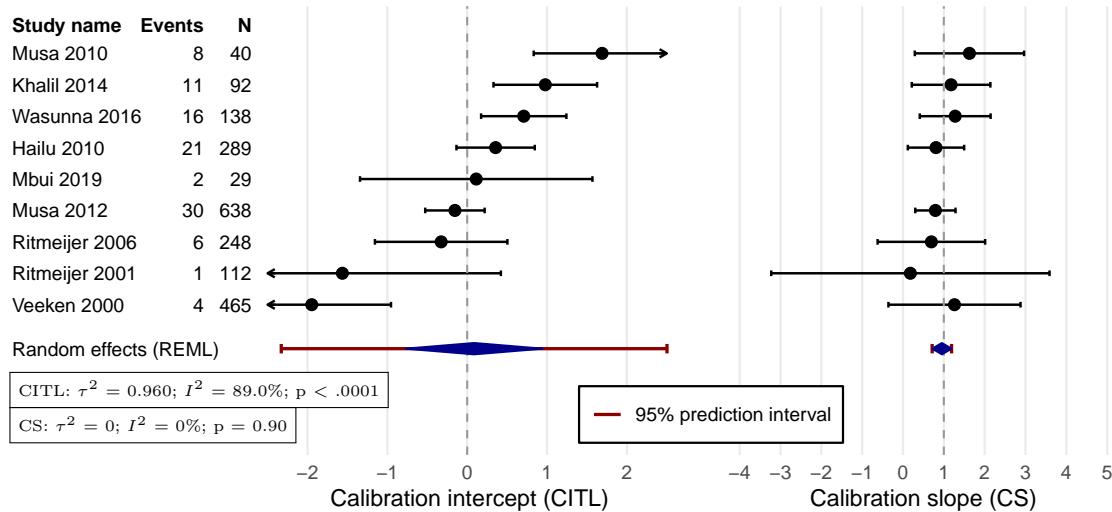


Figure 6.11: Forest plots showing individual and pooled study calibration measures for the model **including** parasite grade. Left: calibration intercept (calibration-in-the-large, CITL); Right: calibration slope (CS). Blue diamonds: summary estimates with 95% confidence intervals. Calibration slope for Mbui 2019 not shown due to small sample size causing failure of model convergence.

with higher relapse risk, and two additional clinical predictors were also selected: spleen size and duration of fever, both showing inverse associations with relapse risk.

Discriminative performance was modest overall and better when parasite grade was available. The pooled apparent c-statistics were 0.72 (95% CI 0.63–0.82) for the model including parasite grade and 0.66 (95% CI 0.56–0.76) for the model excluding it, reducing after optimism correction to 0.68 and 0.61, respectively. Calibration was broadly concordant between observed and predicted relapse risk across predictors, although substantial between-study variation in calibration intercepts (calibration-in-the-large) was observed for both models, alongside moderate-to-high optimism in calibration slope.

In the next chapter, the East Africa and ISC models will be explored through a more critical lens. Reflections will be made on both the clinical plausibility of the identified predictor-outcome relationships and on the practical considerations of model implementation in routine conditions, including how they can be used to improve clinical decision making in the context of elimination.

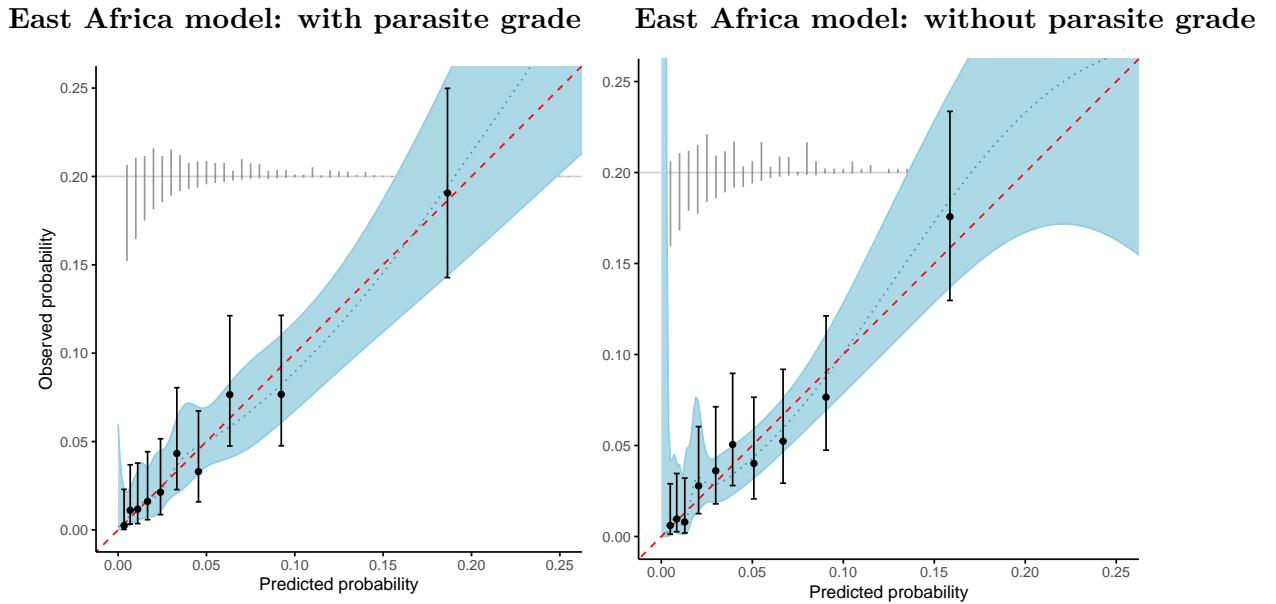


Figure 6.12: Calibration plots showing observed versus predicted probabilities for deciles of predicted probability. Red dashed line represents perfect calibration. Observed probabilities are presented with 95% confidence intervals (black error bars). A generalised additive model is fitted to show the smoothed mean observed probability (blue dotted line) with 95% confidence intervals (blue ribbon). Histograms, normalised by outcome, are overlaid to illustrate the distribution of relapses and cures across the expected probabilities.

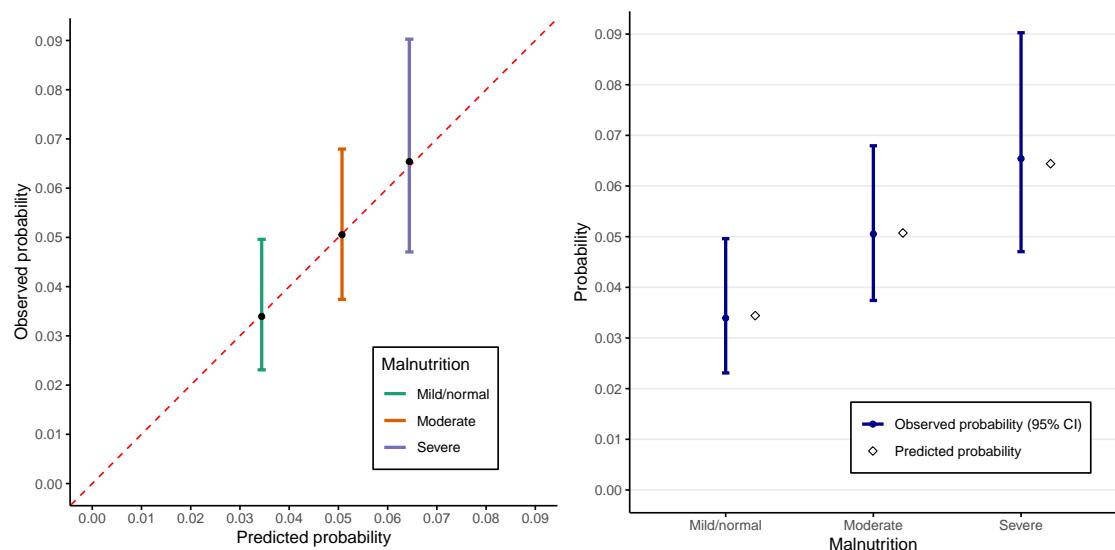


Figure 6.13: Calibration plots for malnutrition severity (model including parasite grade).

7

Discussion

Contents

7.1	Introduction	126
7.2	Interpretation of Predictors and Causal Plausibility	127
7.2.1	Indian Subcontinent	127
7.2.2	East Africa	127
7.3	Implementation Considerations	127
7.3.1	Indian subcontinent	127
7.3.2	East Africa	127
7.4	Strengths	127
7.5	Limitations	127
7.6	Future Research	128
7.7	Conclusion	128

7.1 Introduction

- Brief recap of:
 - The four models
 - The systematic review
 - High-level summary of key findings:
 - Predictive performance
 - Consistently important predictors
 - Regional contrasts
- Explicit statement of discussion aims and structure
- Interpretation
- Clinical relevance
- Implementation feasibility

7.2 Interpretation of Predictors and Causal Plausibility

7.2.1 Indian Subcontinent

7.2.2 East Africa

- Predictors consistently selected across regions • Predictors that diverge by geography
 - Biological and clinical plausibility: • Host factors • Treatment-related factors
 - Immunological hypotheses • Discussion of whether predictors are: • Causal
 - Proxies • Markers of unmeasured processes

7.3 Implementation Considerations

7.3.1 Indian subcontinent

- Model performance and strengths • Key predictors and their regional relevance • Implications for:
 - Follow-up strategies • Targeted surveillance
 - Resource allocation

7.3.2 East Africa

- Differences in predictors and performance • Role of:
 - Comorbidities
 - Treatment heterogeneity
 - Programmatic constraints
 - Implications for clinical and public health practice
- 5. Clinical and programmatic implementation considerations
 - (Strongly recommend keeping this separate) • Who would use these models?
 - At what point in the care pathway?
 - Data availability in routine care
 - Risks of misclassification
 - Ethical considerations:
 - Risk stratification
 - Stigma or differential care
 - Alignment with VL elimination goals

This section is especially valuable for examiners and policy-minded readers.

7.4 Strengths

7.5 Limitations

Your list is good; consider adding framing:

- Data limitations
- Trial populations
- Missingness and measurement error
- Methodological limitations
- Overfitting

- Outcome definition • Lack of time-to-event modelling • Conceptual limitations
- Prediction vs causation • Immune mechanisms inferred but not measured
 - Explicitly note which limitations are unlikely vs likely to materially affect conclusions.

7.6 Future Research

Strong already. You could sharpen by grouping:

- Methodological
- Dynamic prediction models
- Time-to-relapse modelling
- Incorporation of biomarkers
- Biological
- Immune correlates of relapse
- Implementation
- Prospective validation
- Impact studies
- Stakeholder-driven refinement

7.7 Conclusion

- One-two paragraph synthesis of:
 - What is now known that wasn't before
 - Why this matters for VL control
 - Explicit statement of contribution:
 - Scientific
 - Clinical
 - Methodological

Appendices

A

Appendix — Background

Contents

A.1 Background Literature Search	130
---	------------

A.1 Background Literature Search

A literature search was performed with the aim of writing a narrative review on the VL relapse. Specifically to address the following questions:

- What is the burden of relapse?
- When does relapse occur?
- What is the cause (mechanism) of relapse?
- What are the determinants of relapse?

All searches performed on August 11th 2025 on PubMed, Embase and Web of Science (Boxes A.1–A.3). Search performed from database inception. After deduplication with Covidence[180] 1,891 articles were identified. Following title and abstract review, 55 articles were identified for inclusion in the narrative review.

Inclusion criteria:

Inclusion criteria: (1) Publications looking at associations between VL relapse occurrence and host/parasite characteristics during, before, or shortly after treatment of index case. (2) Systematic/literature reviews looking at relapse determinants. (3) All ages, geographical locations.

Exclusion criteria: (1) Conference abstracts, protocols, case reports. (2) Articles not published in English. (3) Non-human studies.

Box A.1: PubMed search terms (932 hits)

```
("Leishmaniasis, Visceral"[Mesh] OR "visceral leishmaniasis" OR  
"leishmaniasis, visceral" OR "kala azar" OR "kala-azar") AND  
("Recurrence"[Mesh] OR relapse* OR recurrent OR recurrence OR  
recrudescence OR "treatment failure")
```

Box A.2: Embase search terms (1303 hits)

1. visceral leishmaniasis.mp. or exp visceral leishmaniasis/
2. kala-azar.mp.
3. 1 or 2
4. exp relapse/ or relapse*.mp.
5. recurrence*.mp. or exp recurrent disease/
6. recurrent.mp.
7. recrudescence.mp.
8. 4 or 5 or 6 or 7
9. treatment failure.mp. or exp treatment failure/
10. 8 or 9
11. 3 and 10

Box A.3: Web of Science (1097 hits)

1. TS=((Leishmaniasis and Visceral) or (Kala azar) or (Kala azar))
2. TS=((relapse\$) or (recurren\$) or (recrudescence) or (treatment failure\$))
3. #2 AND #1

B

Appendix — Model methodology

C

Appendix — Results: Indian Subcontinent Model

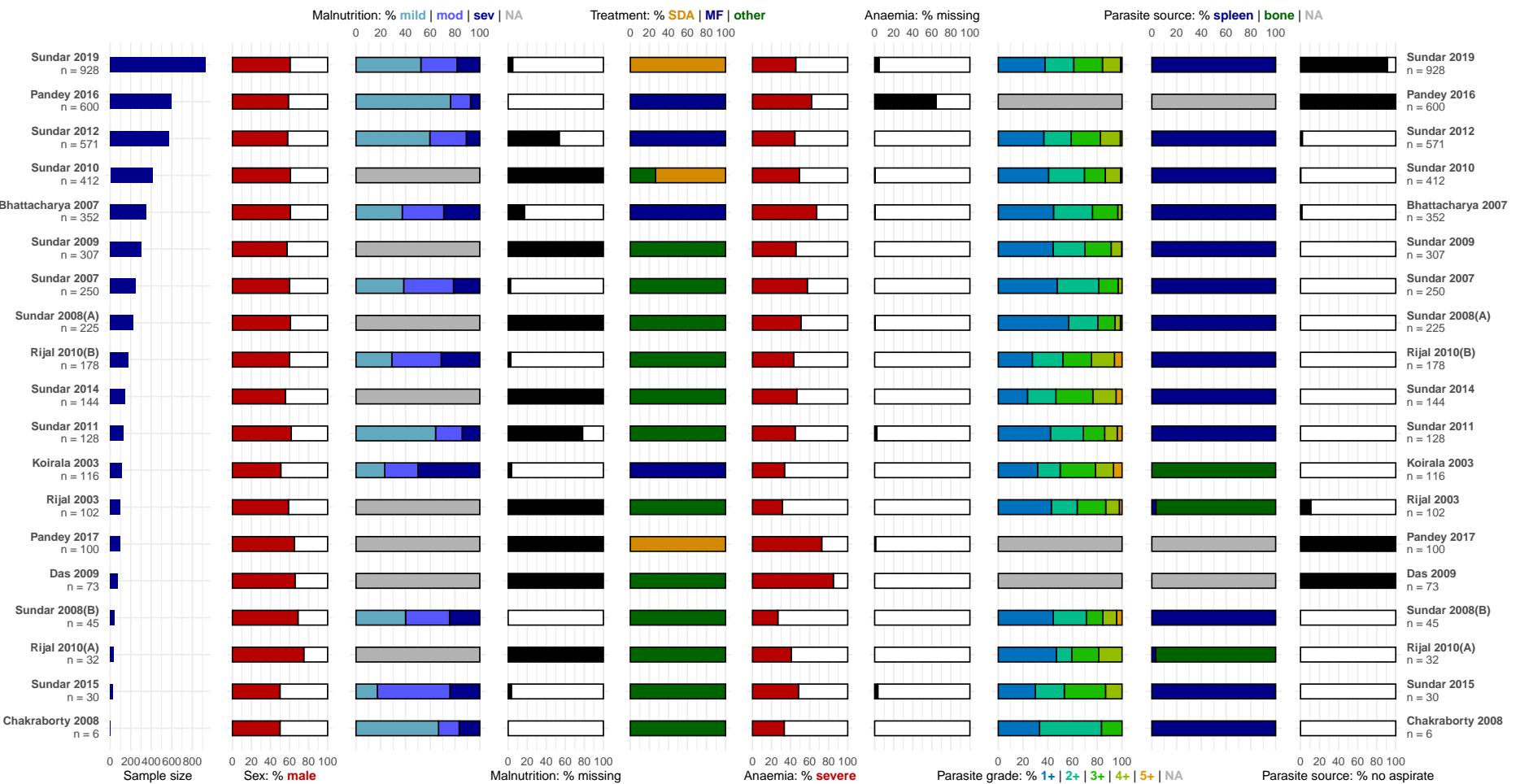


Figure C.1: Distribution of categorical predictors by study. Missing data excluded from stacked bar charts for malnutrition, parasite grade. Where aspirates were performed, the source was never missing. NA: unable to show distribution as all study-specific data are missing. MF: miltefosine; SDA: single dose liposomal amphotericin B

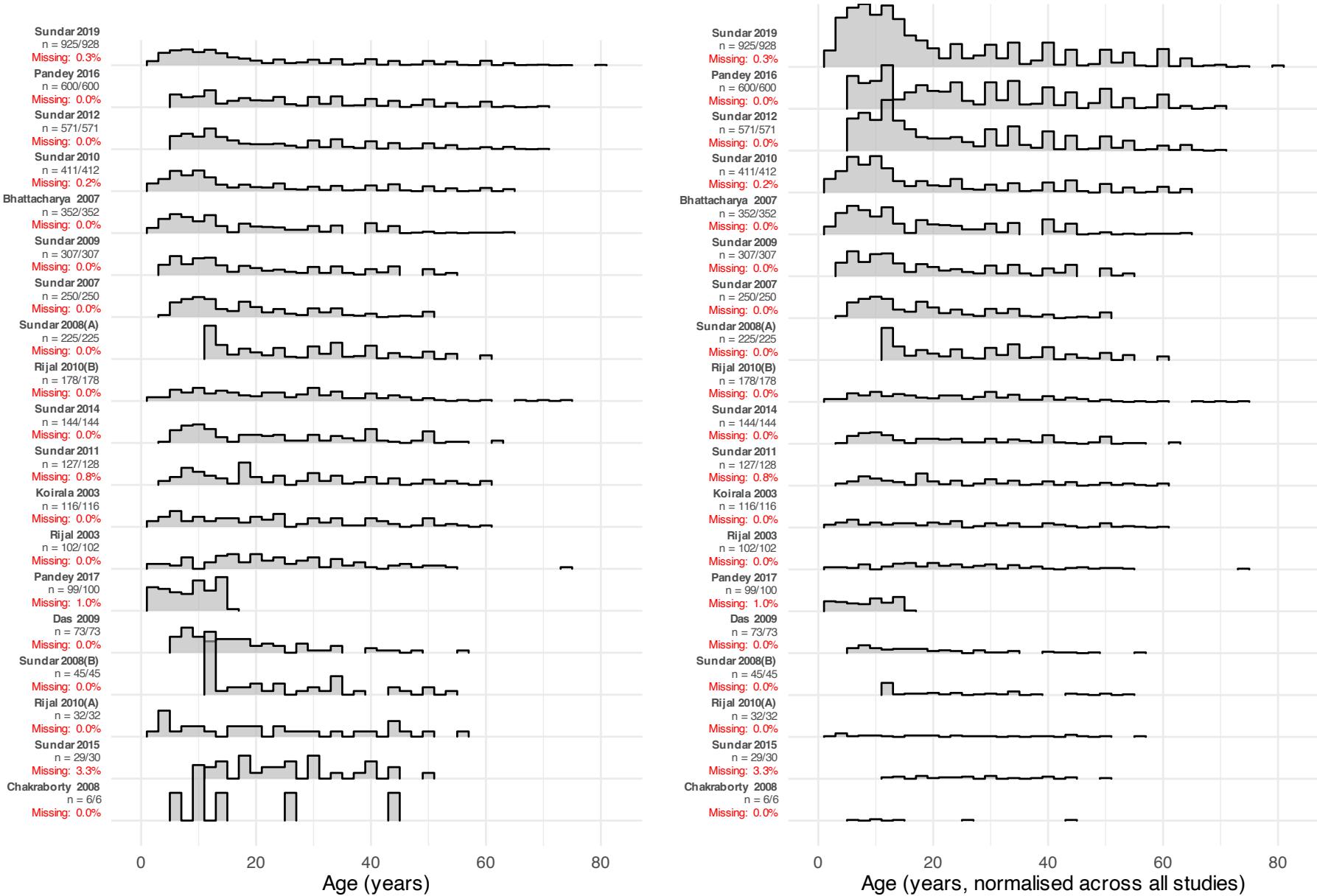


Figure C.2: Distribution of age across studies from the Indian subcontinent. Missing age data described by study.

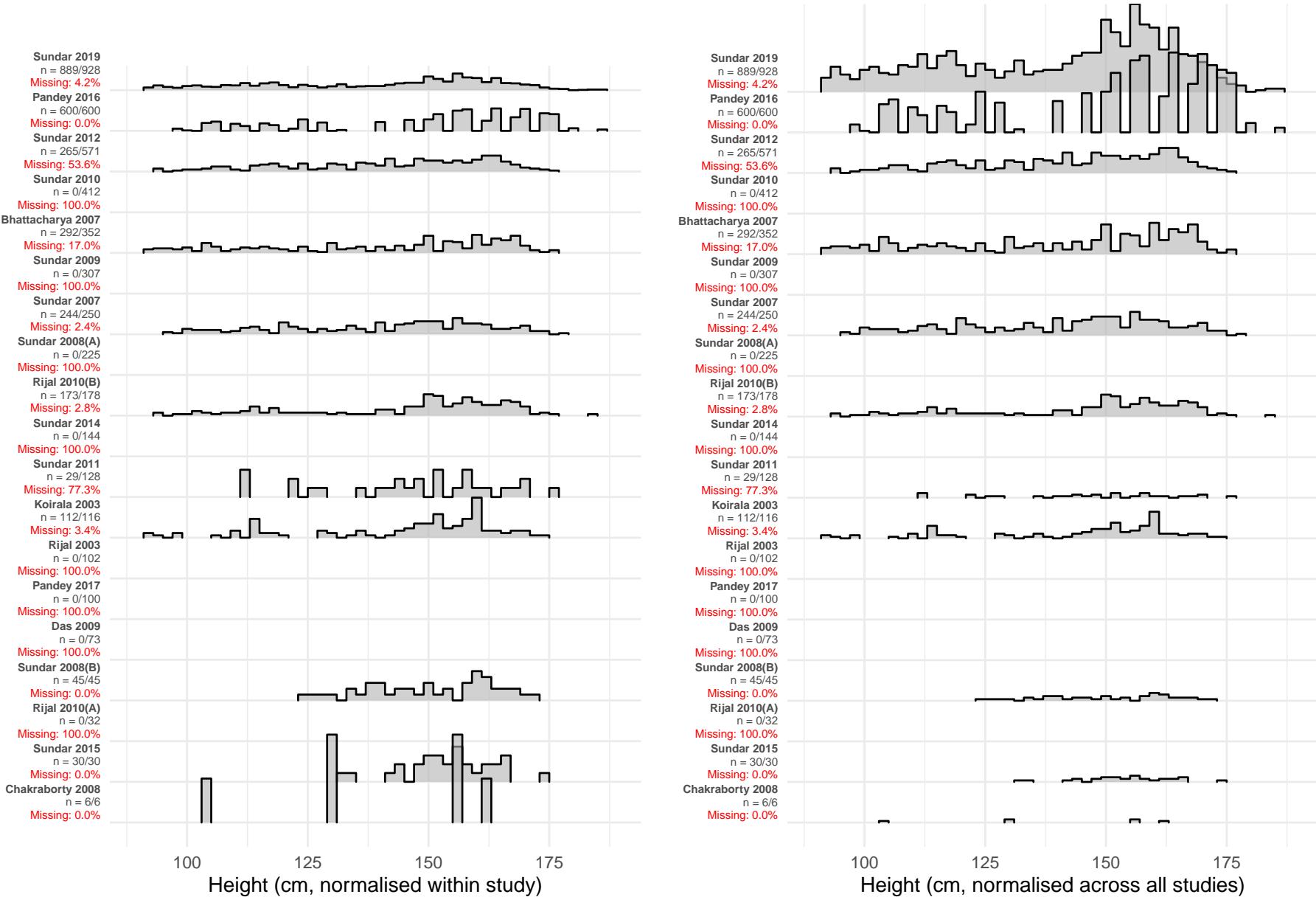


Figure C.3: Distribution of height across studies from the Indian subcontinent. Missing data described by study.

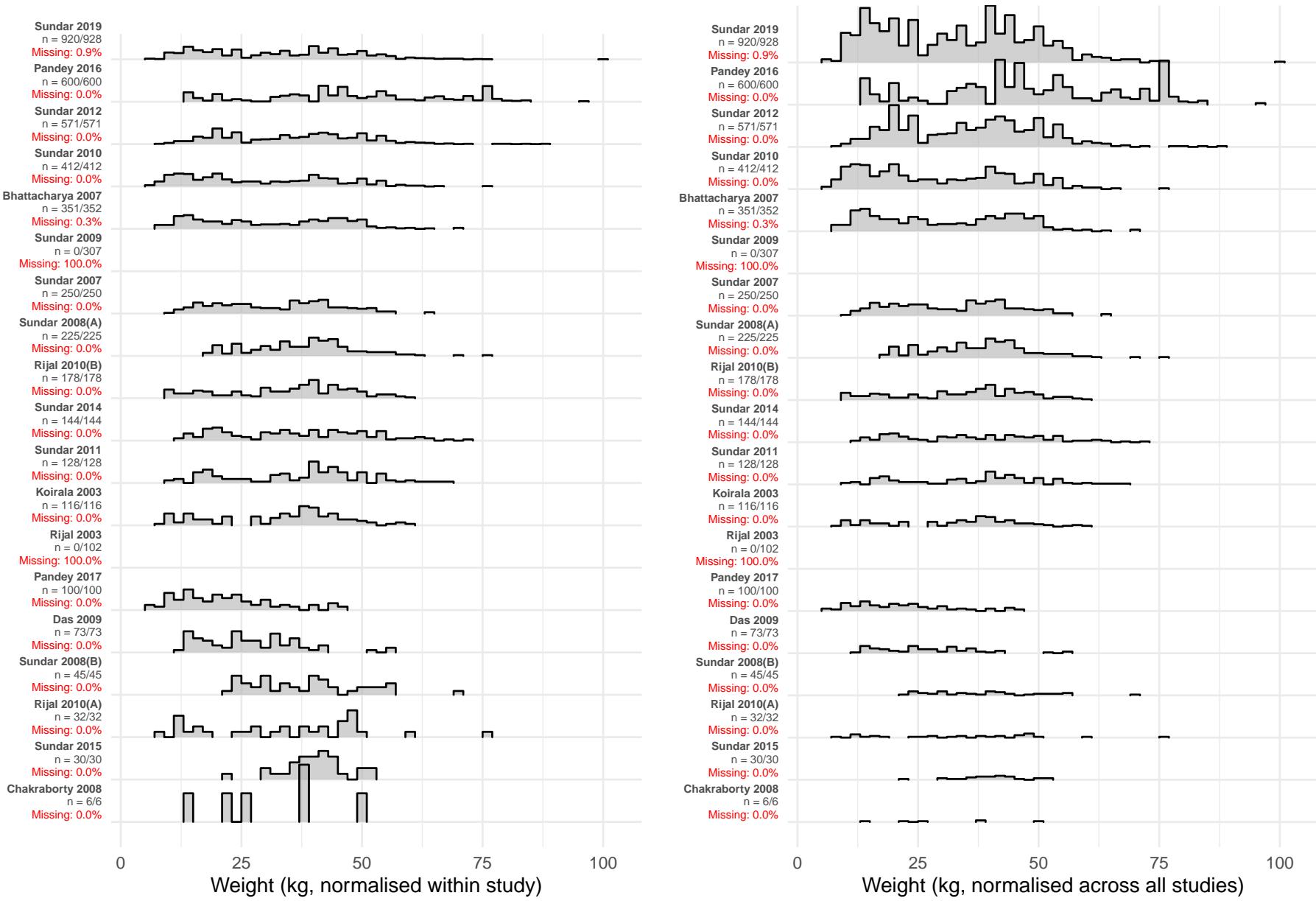


Figure C.4: Distribution of weight across studies from the Indian subcontinent. Missing data described by study.

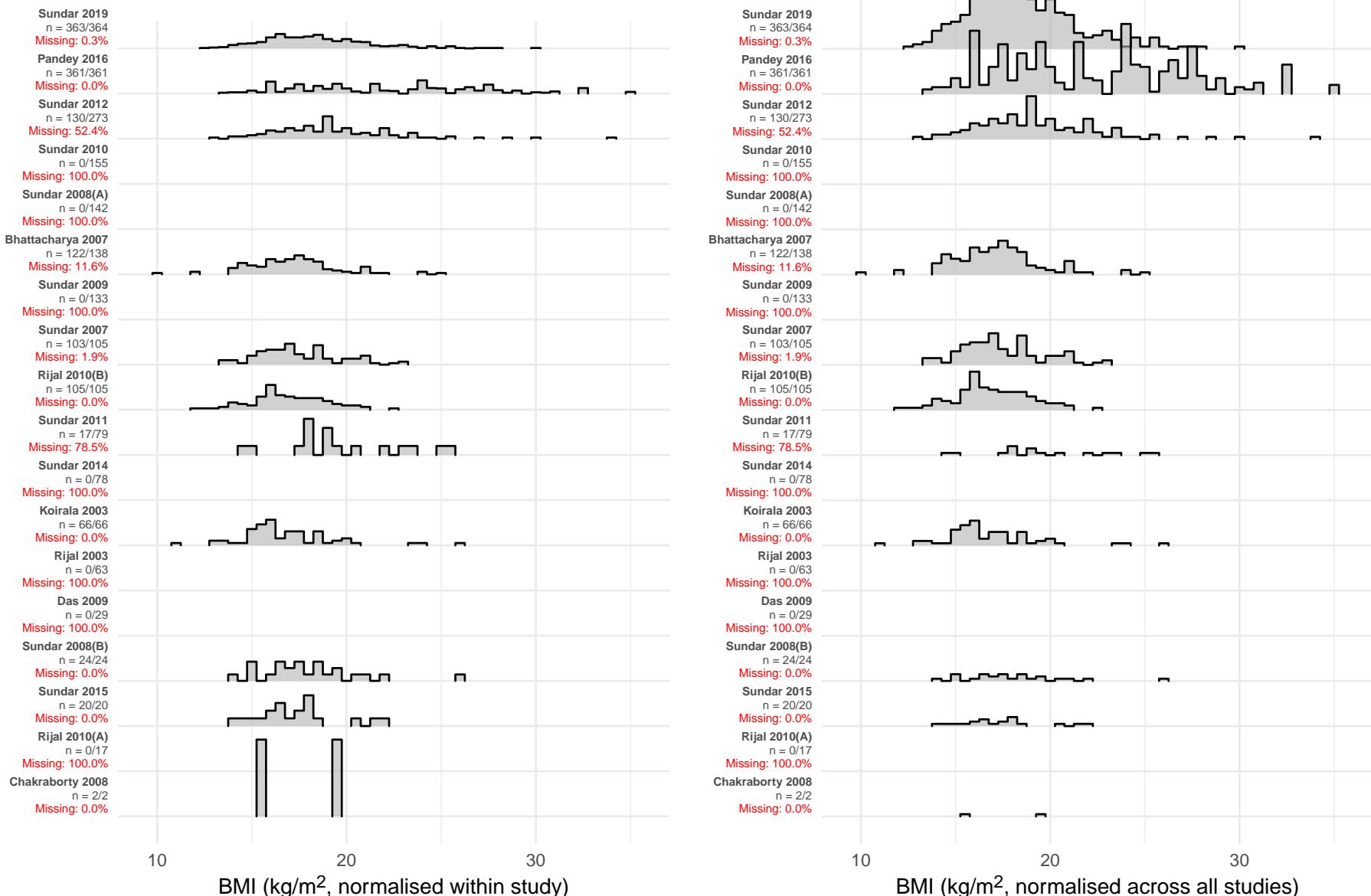


Figure C.5: Distribution of BMI across studies from the Indian subcontinent. Including only participants aged 19 and over (Pandey 2017 excluded). Missing data described by study.

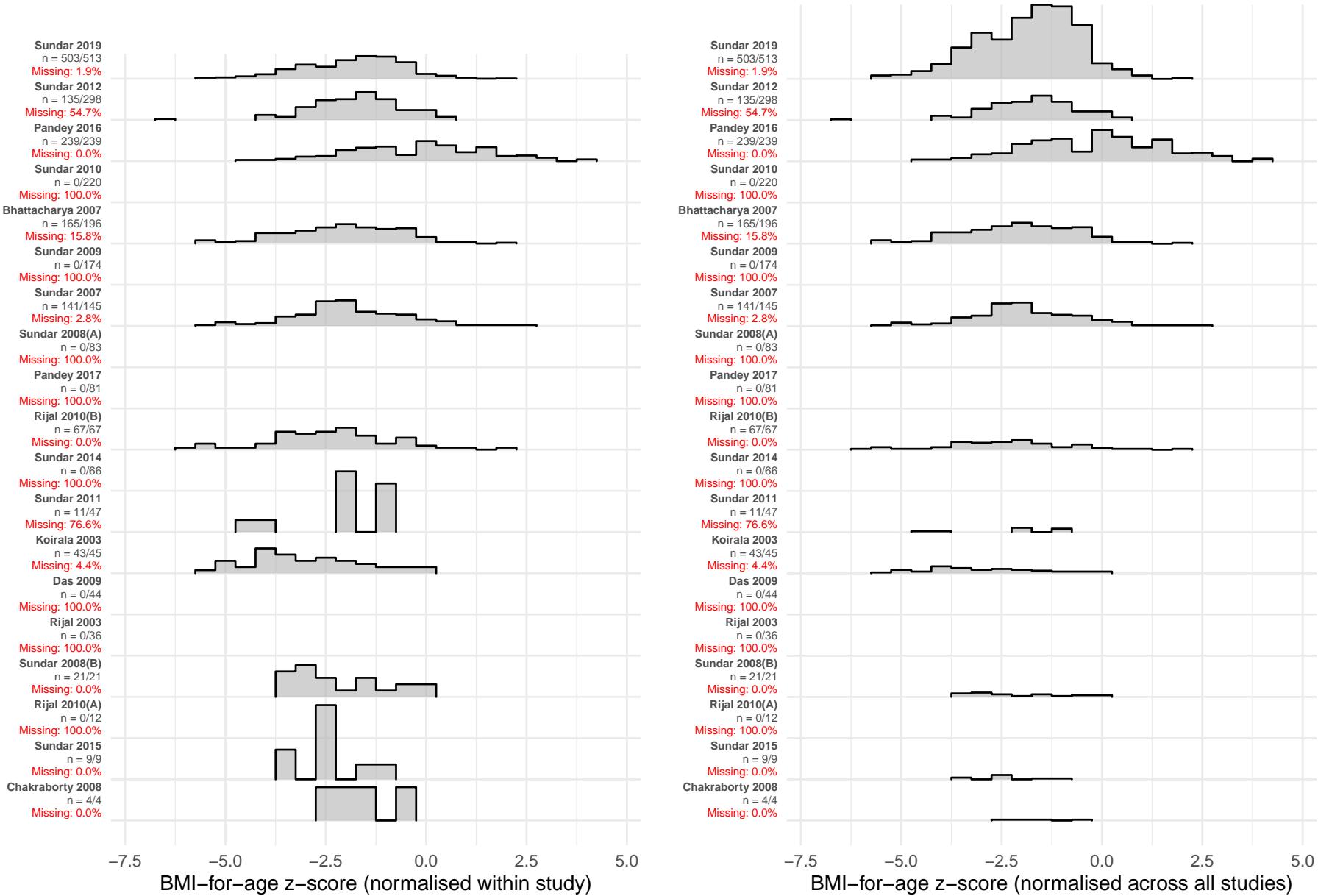


Figure C.6: Distribution of BMI-for-age z score across studies from the Indian subcontinent. Including only participants aged from 5–18, inclusive. Missing data described by study.

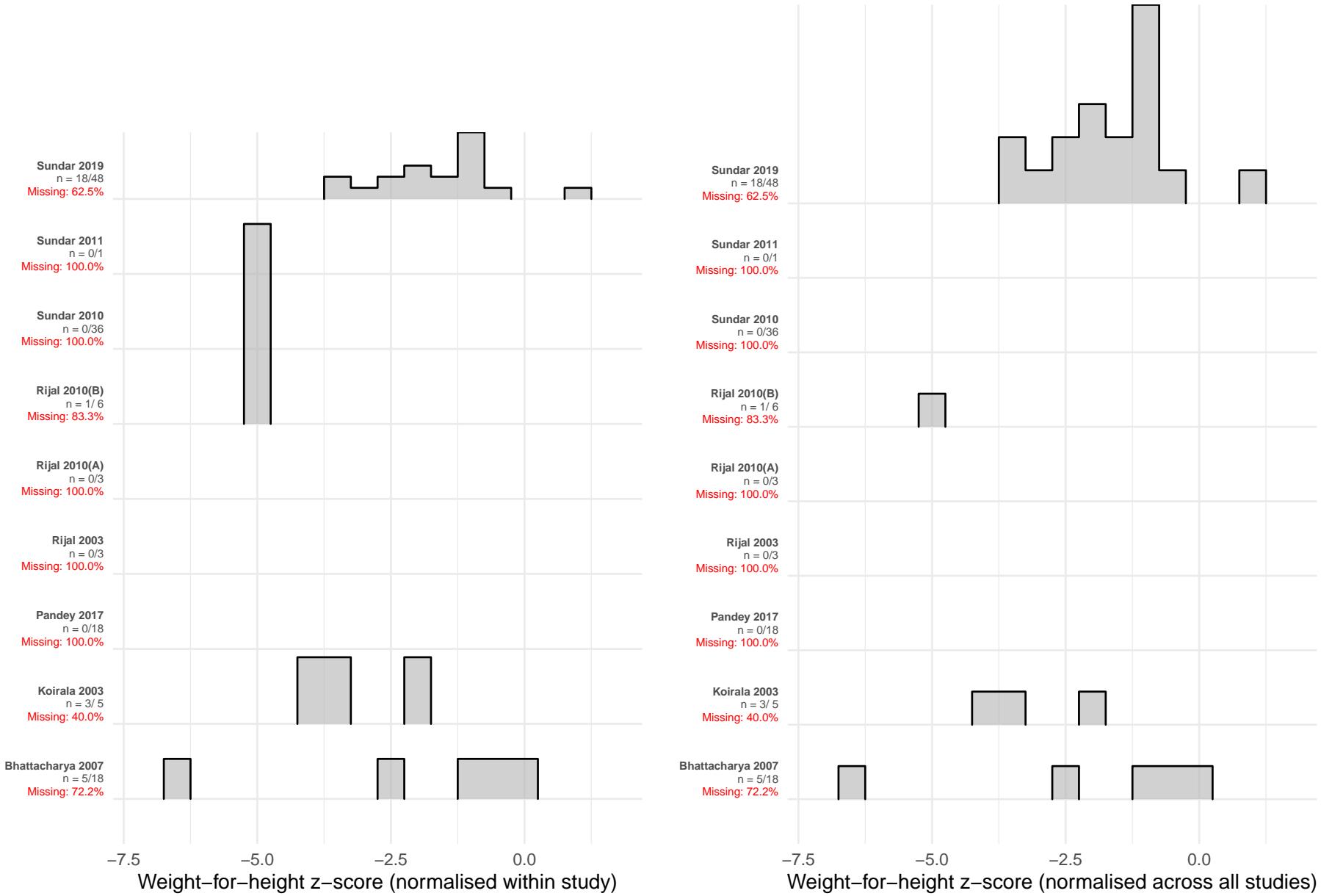


Figure C.7: Distribution of weight-for-height z score across studies from the Indian subcontinent. Including only participants aged under 5. Missing data described by study.

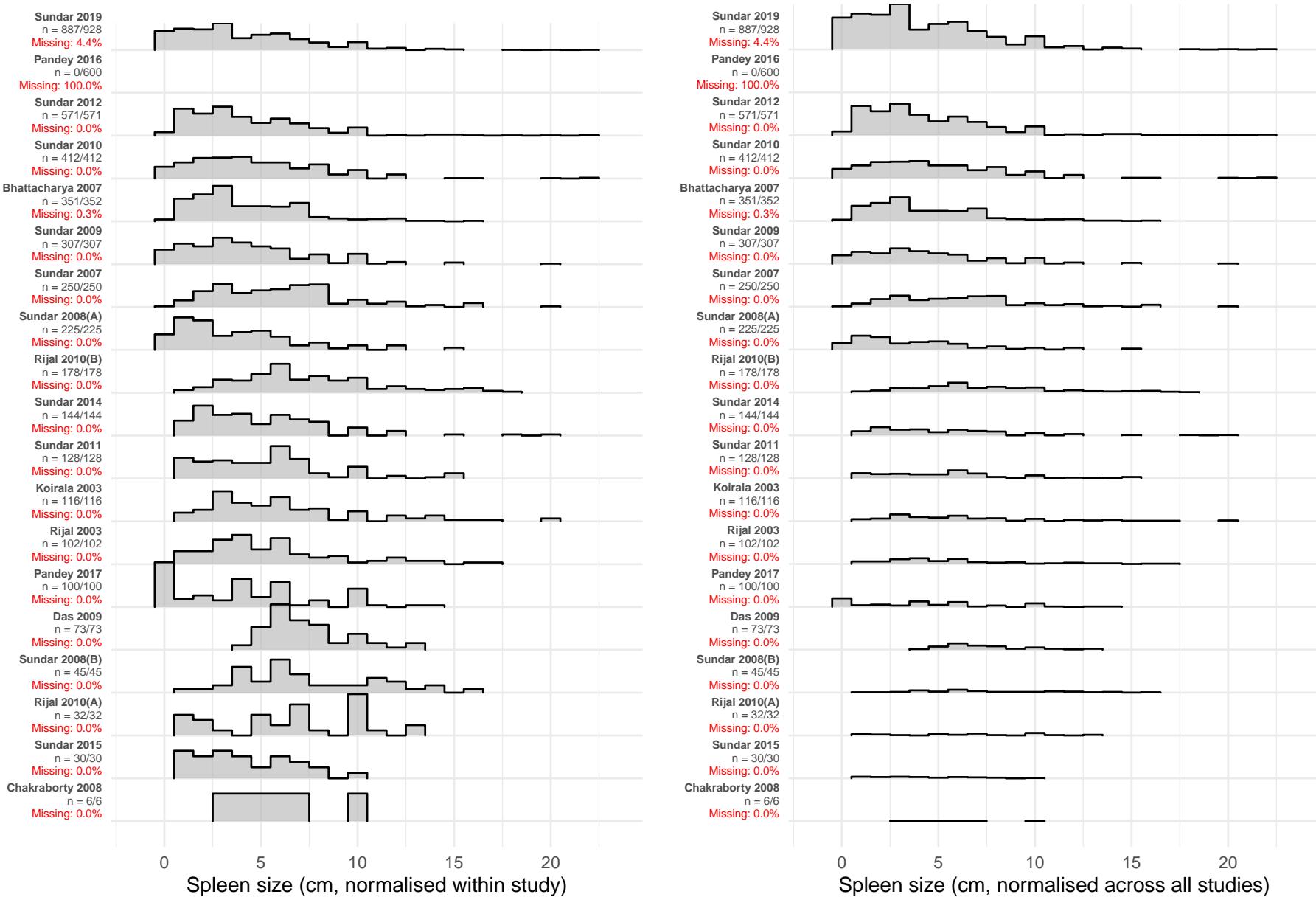


Figure C.8: Distribution of spleen size across studies from the Indian subcontinent. Missing data described by study.

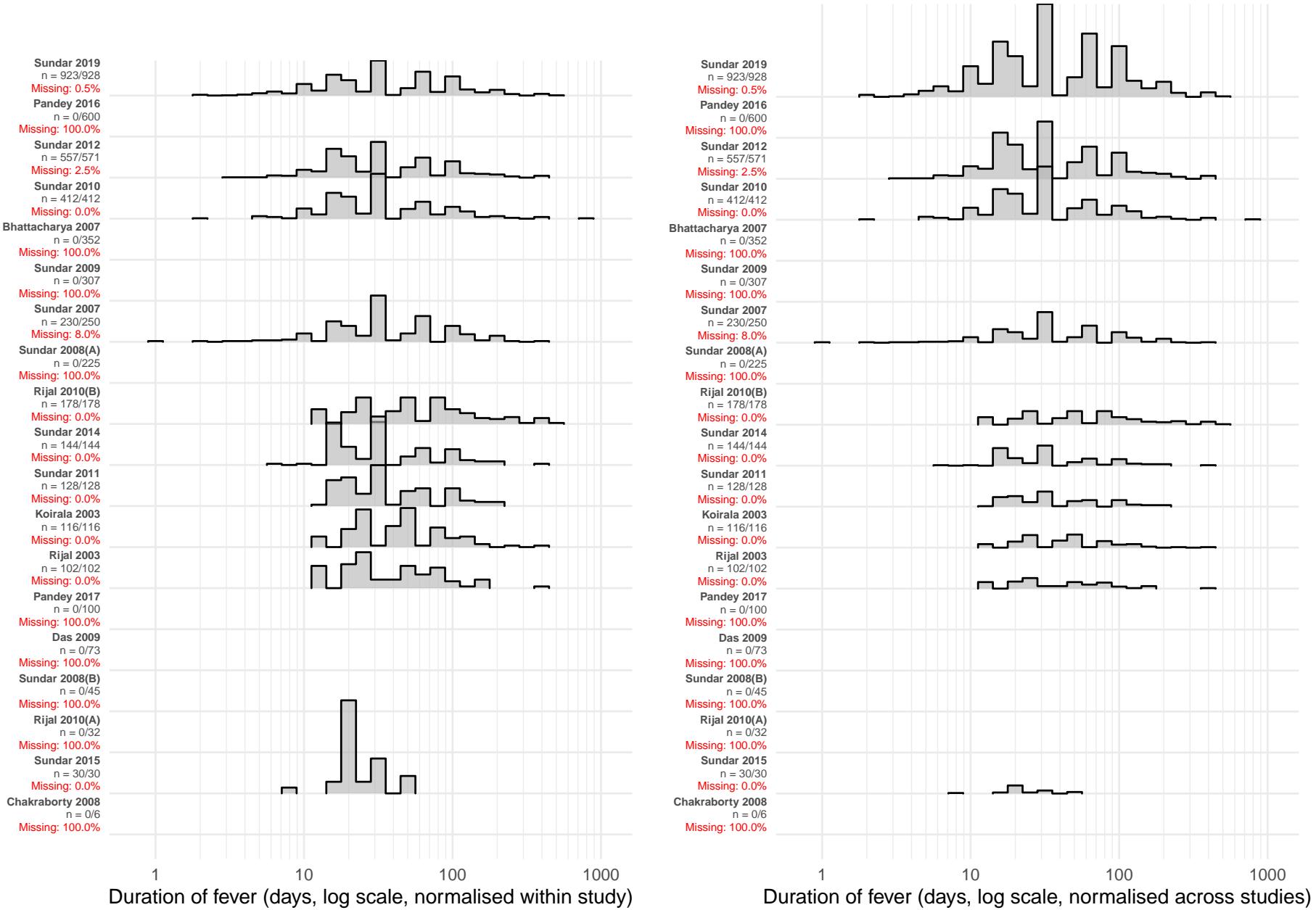


Figure C.9: Distribution of fever duration (log scale) across studies from the Indian subcontinent. Missing data described by study.

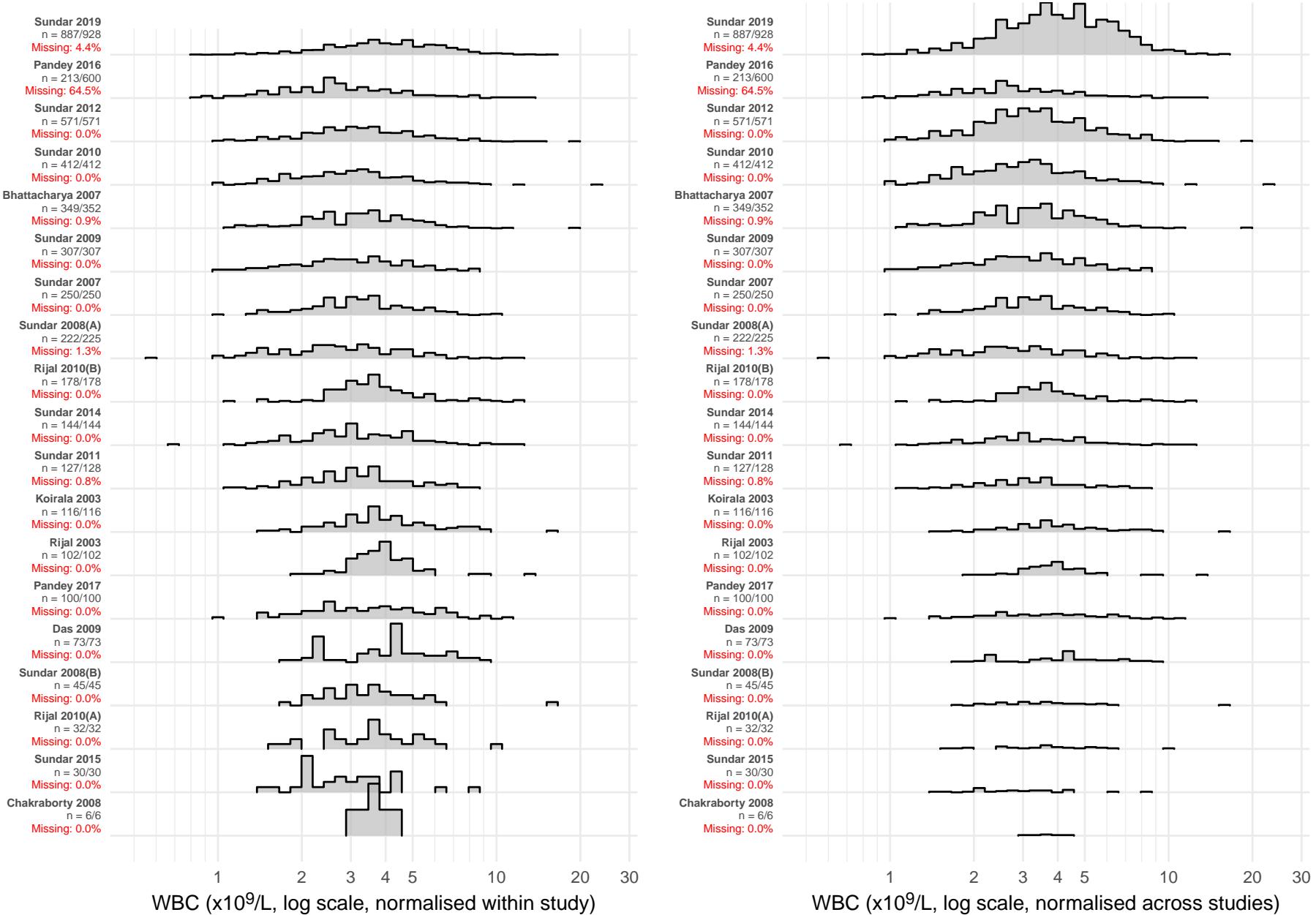


Figure C.10: Distribution of white blood cell count (log scale) across studies from the Indian subcontinent. Missing data described by study.

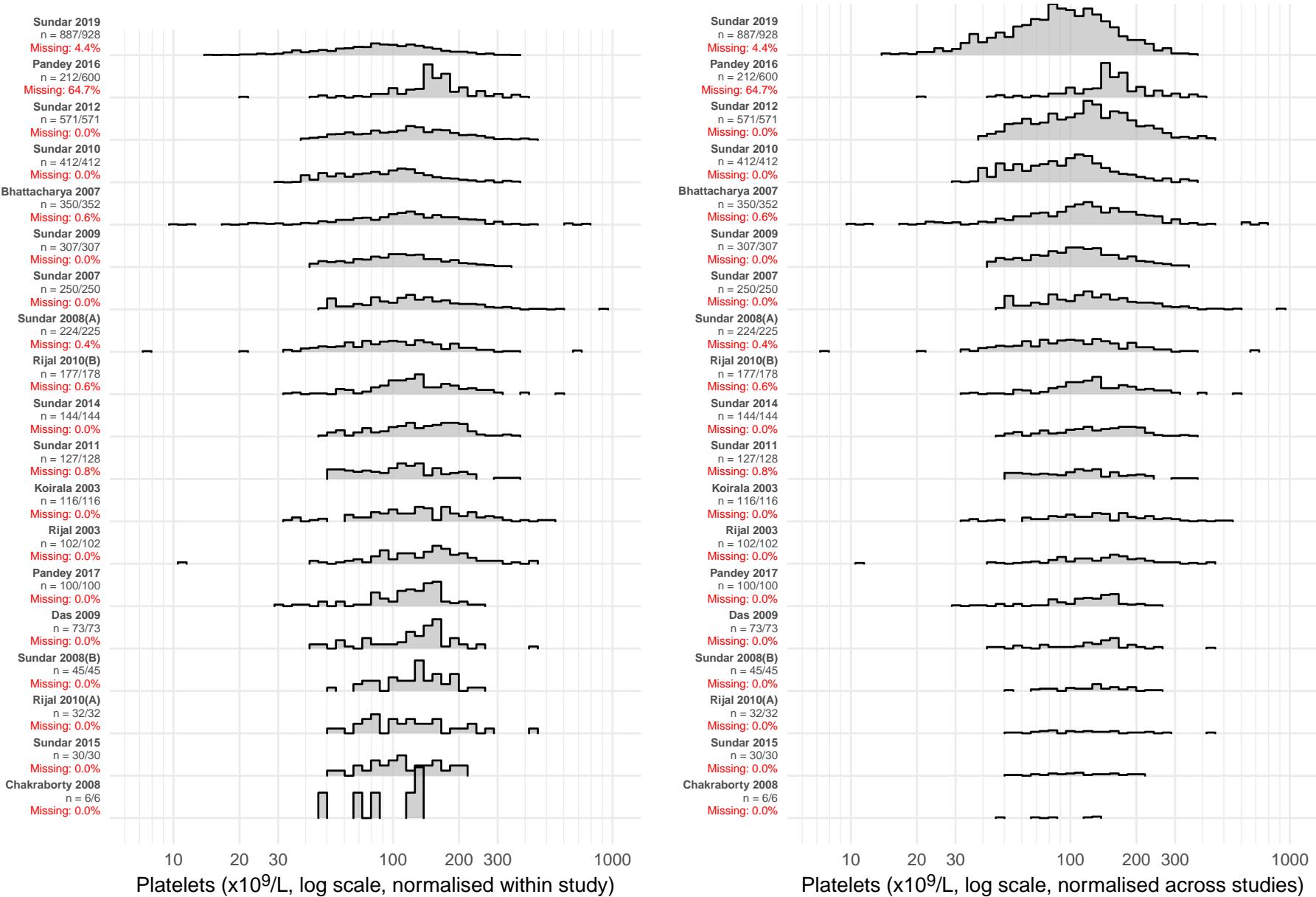


Figure C.11: Distribution of platelet count (log scale) across studies from the Indian subcontinent. Missing data described by study.

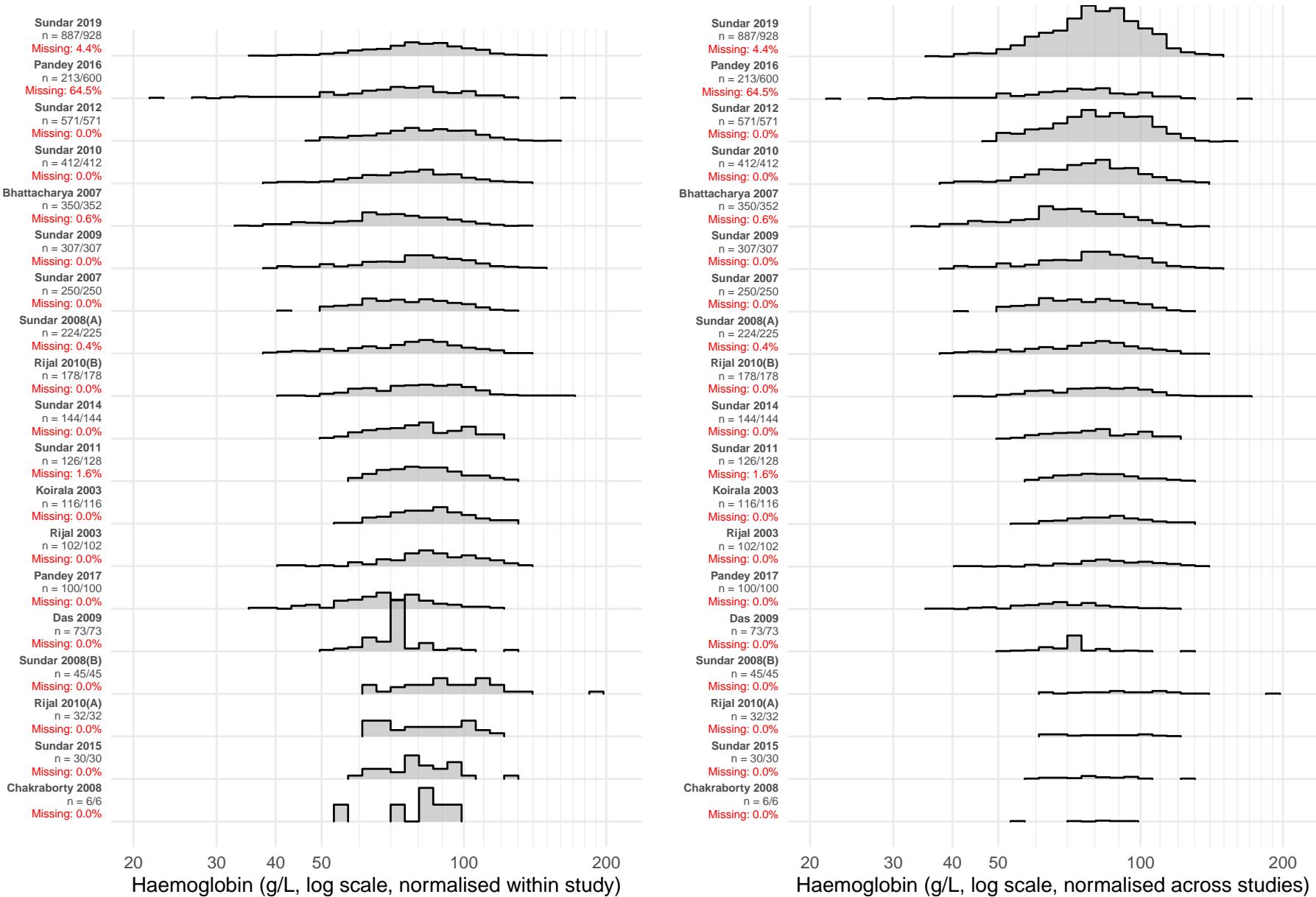


Figure C.12: Distribution of haemoglobin (log scale) across studies from the Indian subcontinent. Missing data described by study.

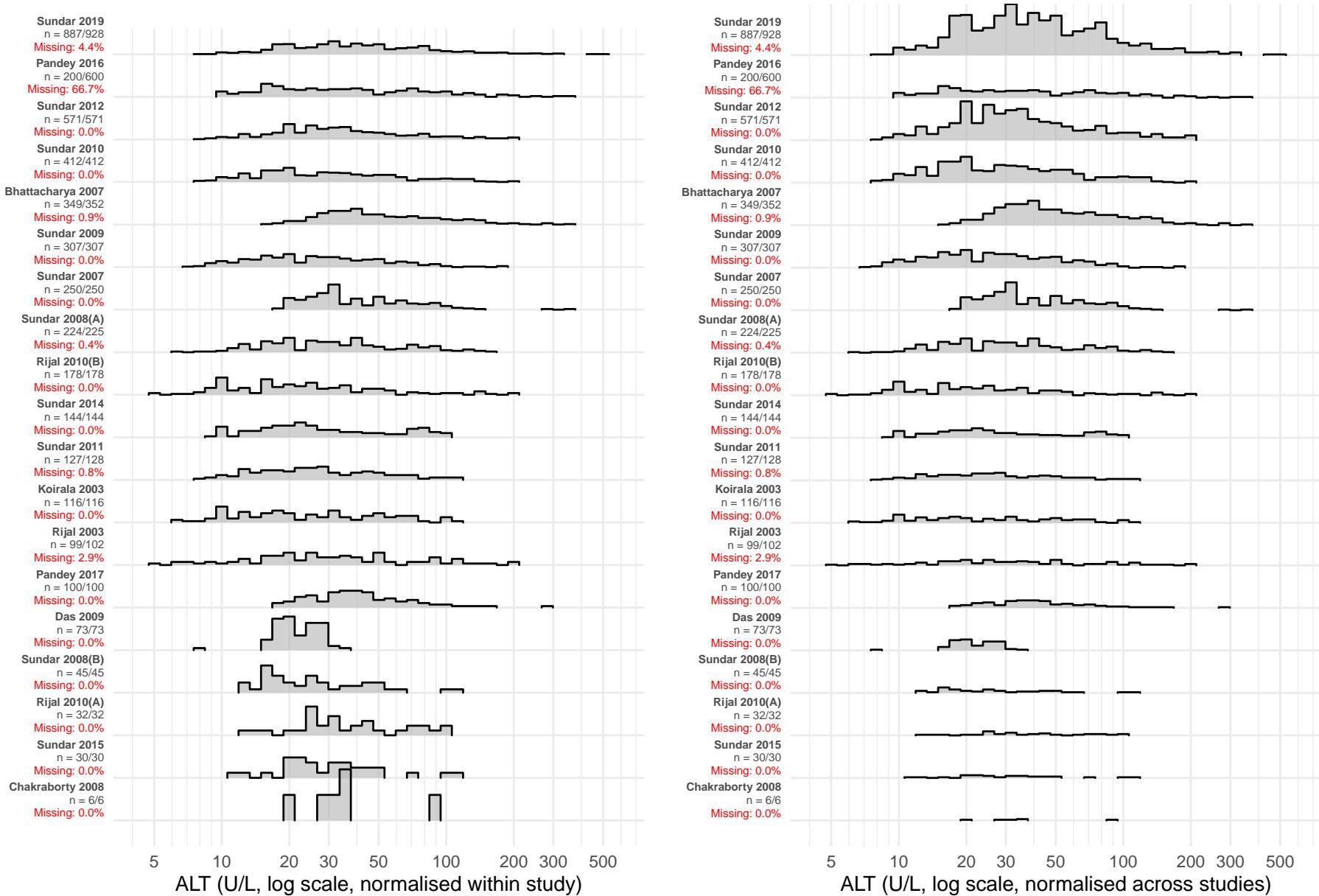


Figure C.13: Distribution of alanine transaminase (ALT, log scale) across studies from the Indian subcontinent. Missing data described by study.

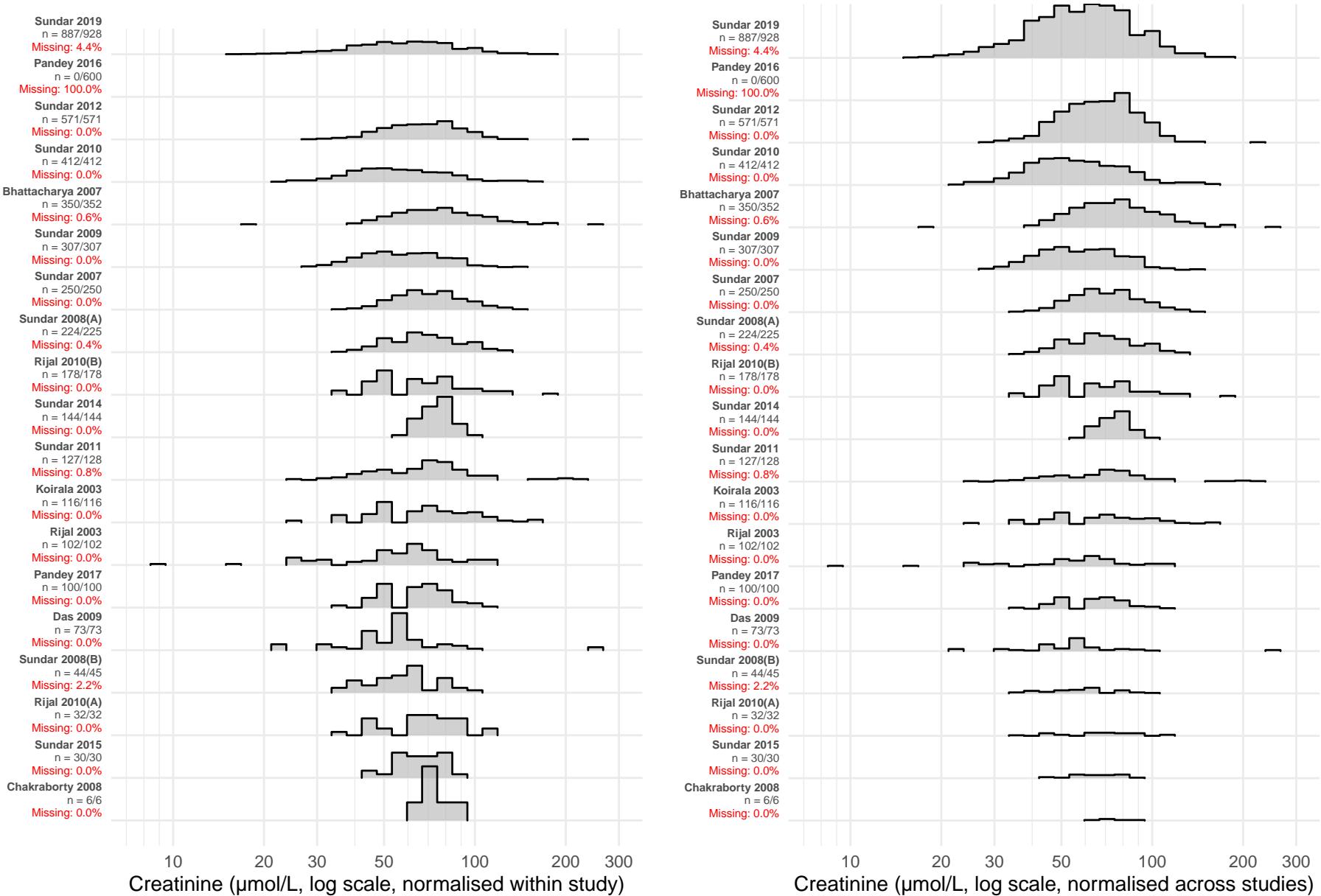


Figure C.14: Distribution of creatinine (log scale) across studies from the Indian subcontinent. Missing data described by study.

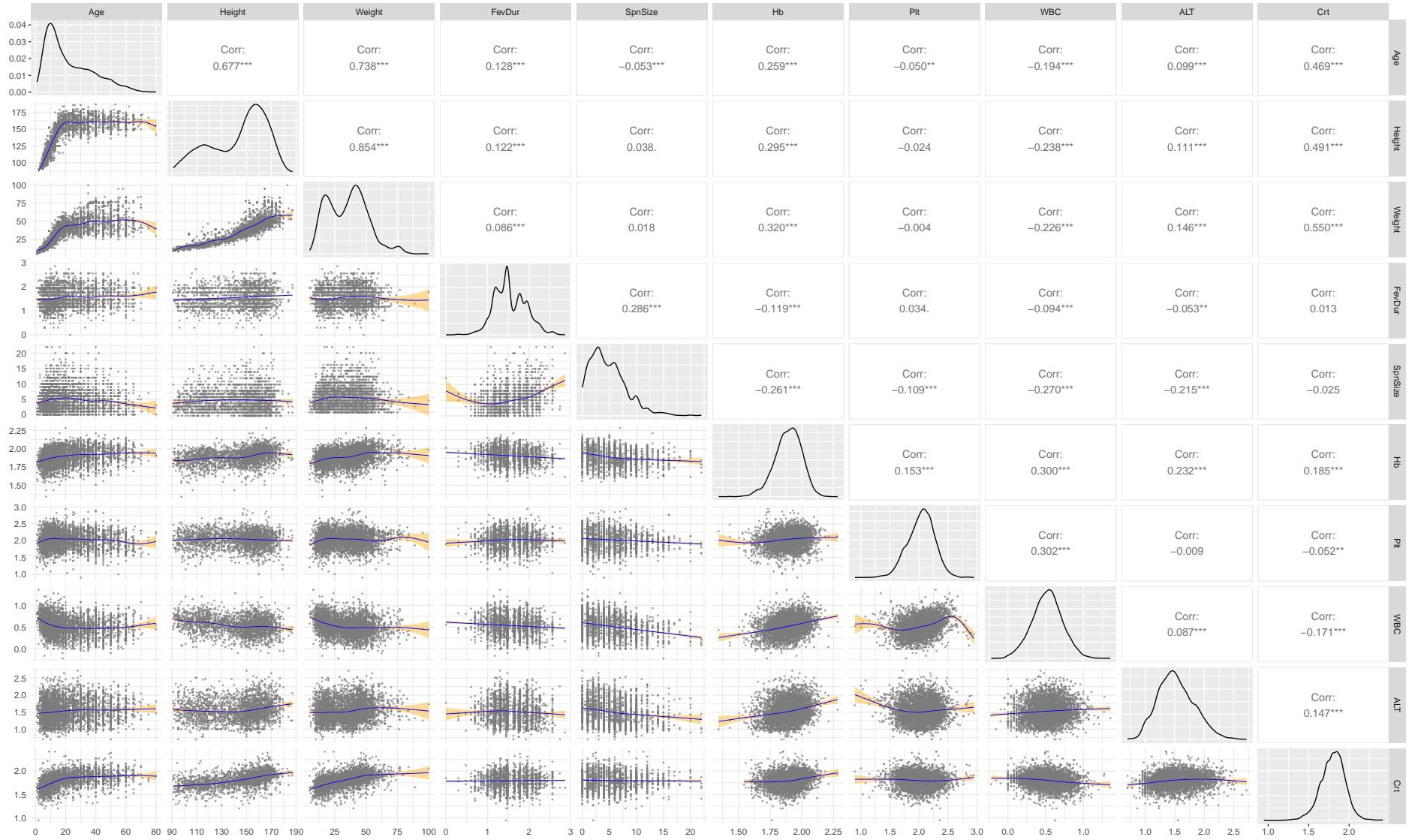


Figure C.15: Correlation between continuous variables. For scatter plots, a univariable generalised additive model is fitted (blue line) with 95% confidence interval ribbon filled (orange area). Pearson correlation coefficients are presented, *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$, . $p < 0.10$. Age: years; Height: cm; Weight: kg; FevDur: duration of fever, log10(days); SpnSize: spleen size (cm); Hb: haemoglobin, log10(g/L); Plt: platelets, log10($\times 10^9/L$); WBC: white blood cells log10($\times 10^9/L$); ALT: alanine aminotransferase, log10(U/L); Cr: creatinine log10($\mu\text{mol}/L$).

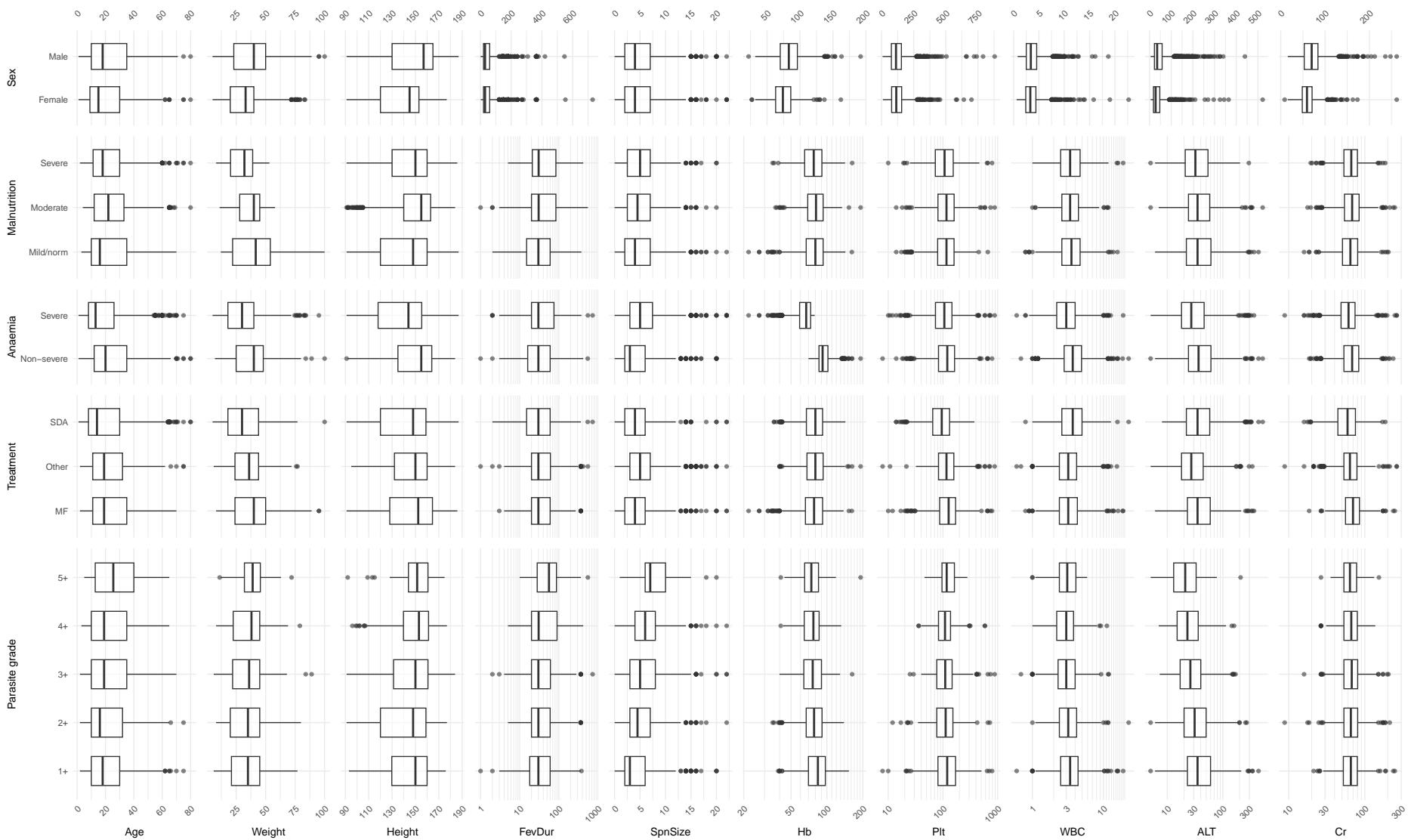


Figure C.16: Correlation between continuous and categorical variables. FevDur and laboratory tests axes transformed to log10 scale. Age: years; Height: cm; Weight: kg; FevDur: duration of fever, days; SpnSize: spleen size, cm; Hb: haemoglobin, g/L; Plt: platelets, $\times 10^9/L$; WBC: white blood cells $\times 10^9/L$; ALT: alanine aminotransferase, U/L; Cr: creatinine $\mu\text{mol}/L$.

term	estimate ¹	std.error	t statistic	df (effective)	p value
Fixed intercept	-2.9166	0.2592	-11.25	1123.60	6.54E-28
Parasite grade ²	0.2669	0.0781	3.42	174.73	7.88E-04
Age (linear) ³	-0.2352	0.0975	-2.41	3222.98	1.59E-02
Age (squared) ³	0.1352	0.0658	2.05	4178.35	4.00E-02
Anaemia: Severe ⁴	-0.3860	0.1597	-2.42	875.81	1.59E-02
Fever duration ⁵	-0.5594	0.1175	-4.76	77.95	8.68E-06
Treatment: SDA ⁶	-0.8349	0.3483	-2.40	3059.66	1.66E-02
Treatment: Other ⁶	-0.7159	0.2880	-2.49	2905.85	1.30E-02
Unadjusted intercept (Sundar 2019):	-2.9888				
Uniform shrinkage factor:	0.9132				
Adjusted intercepts after shrinkage: Average:	-3.1608; Sundar 2019:	-3.2514			
Pooled τ^2 :	0.1209.	Pooled ICC:	0.0354		

¹ Logit scale before uniform shrinkage.

² Continuous variable, from 1+ to 6+ as described by Chulay and Bryceson[52]. Transformed $x \mapsto x - 1$.

³ Units: years. Standardised by mean (22.301633) and standard deviation (15.4159649).

⁴ Reference group: non-severe anaemia.

⁵ Units: days. Converted to natural log scale and then standardised by mean (3.550509) and standard deviation (0.8693318).

⁶ Reference group: standard dose miltefosine.

Table C.1: Model coefficients pooled across 30 imputed datasets using Rubin's Rules. Model **including** parasite grade. df: degrees of freedom; SDA: single dose liposomal amphotericin B; std.error: standard error

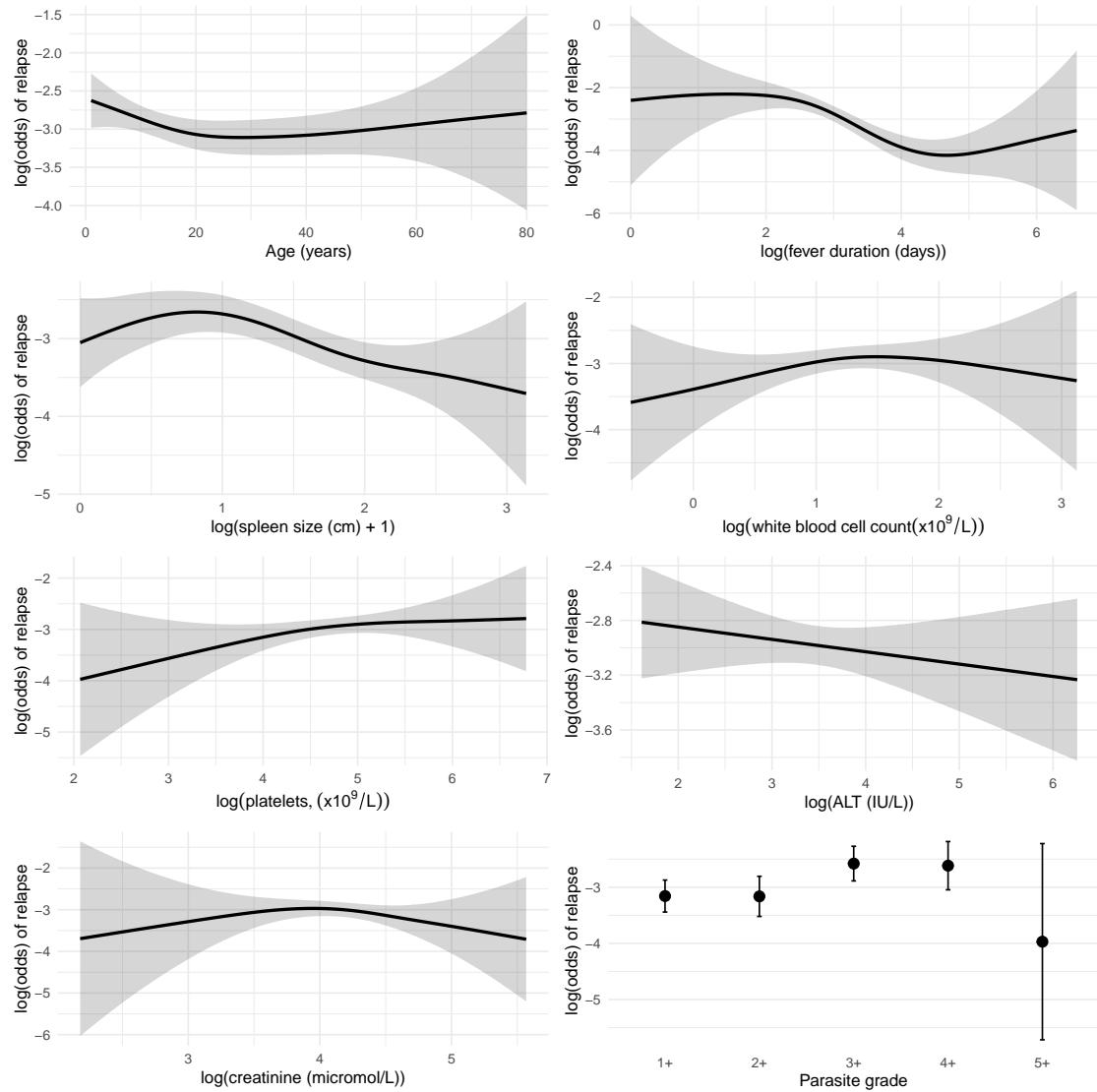


Figure C.17: Associations between transformed continuous predictors and $\log(\text{odds})$ of relapse. For each predictor (excluding parasite grade), a univariable generalised additive model spline fit is shown, with 95% confidence intervals. For parasite grade, 95% confidence intervals are calculated for each grade using the Wilson method.

term	estimate ¹	std.error	t statistic	df	p.value
Fixed intercept	-2.6042	0.2547	-10.22	2137.17	5.51E-24
Age (linear) ²	-0.2161	0.0962	-2.25	3968.52	2.47E-02
Age (squared) ²	0.1434	0.0645	2.22	4413.49	2.62E-02
Anaemia: Severe ³	-0.2921	0.1501	-1.95	2350.48	5.17E-02
Fever duration ³	-0.5221	0.1057	-4.94	107.97	2.86E-06
Treatment: SDA ⁴	-0.8438	0.3696	-2.28	3813.56	2.25E-02
Treatment: Other ⁴	-0.7571	0.3078	-2.46	2506.34	1.40E-02
Unadjusted intercept (Sundar 2019):	-2.7030				
Uniform shrinkage factor:	0.9174				
Adjusted intercepts after shrinkage: Average:	-2.8425; Sundar 2019:	-2.9543			
Pooled τ^2 :	0.1507.	Pooled ICC:	0.0437		

¹ Logit scale before uniform shrinkage.

² Units: years. Standardised by mean (22.301633) and standard deviation (15.4159649).

³ Reference group: non-severe anaemia.

⁴ Units: days. Converted to natural log scale and then standardised by mean (3.550509) and standard deviation (0.8693318).

⁵ Reference group: standard dose miltefosine.

Table C.2: Model coefficients pooled across 30 imputed datasets using Rubin’s Rules. Model **excluding** parasite grade. df: degrees of freedom; SDA: single dose liposomal amphotericin B; std.error: standard error

Predictors	Model with PG	Model without PG
Fever duration	500¹	500¹
Parasite grade	472¹	n/a
Age (linear)	418¹	415¹
Age (squared)	355¹	392¹
Treatment	333¹	337¹
Anaemia	314¹	263¹
Spleen size	244	120
Sex	209	217
Age (cubic)	114	150
ALT	86	184
Platelets	78	78
WBC	51	55
Malnutrition	33	26
Creatinine	30	30

¹ Selected in final model.

Table C.3: Frequency of predictor selection during 500 bootstrap model developments. Predictors sorted by selection frequency in the model with parasite grade. ALT: alanine aminotransferase; n/a: not applicable; PG: parasite grade; WBC: white blood cell count

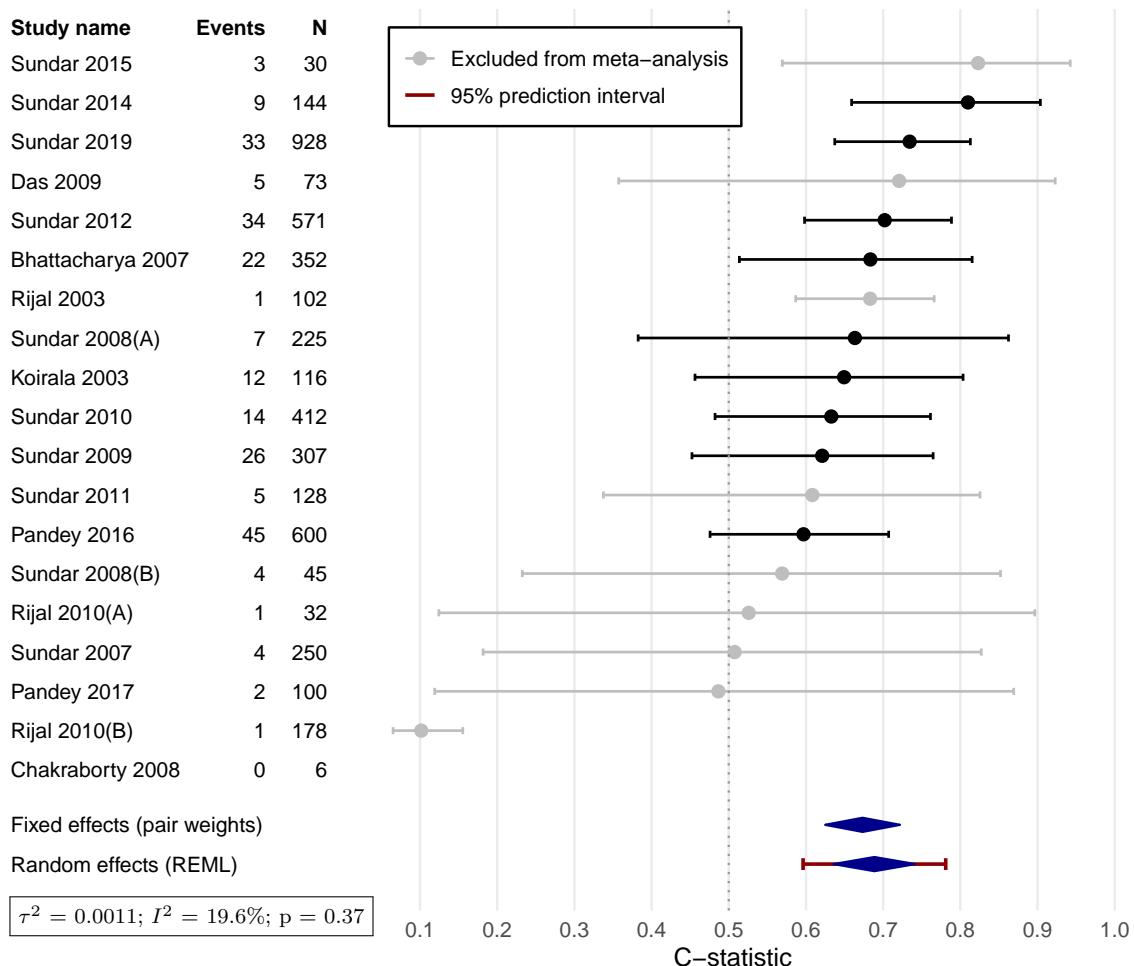


Figure C.18: Forest plot showing individual and pooled study c-statistics for the model **excluding** parasite grade. Pooled c-statistics are presented from both fixed-effects and random-effects meta-analysis models. Blue diamonds: pooled summary estimates with 95% confidence intervals. For Chakraborty 2008, no relapse events occurred and the c-statistic is therefore undefined. Study-specific confidence intervals should be interpreted with caution due to small sample sizes and relapse events in some studies (see Methodology Section 4.3.7).

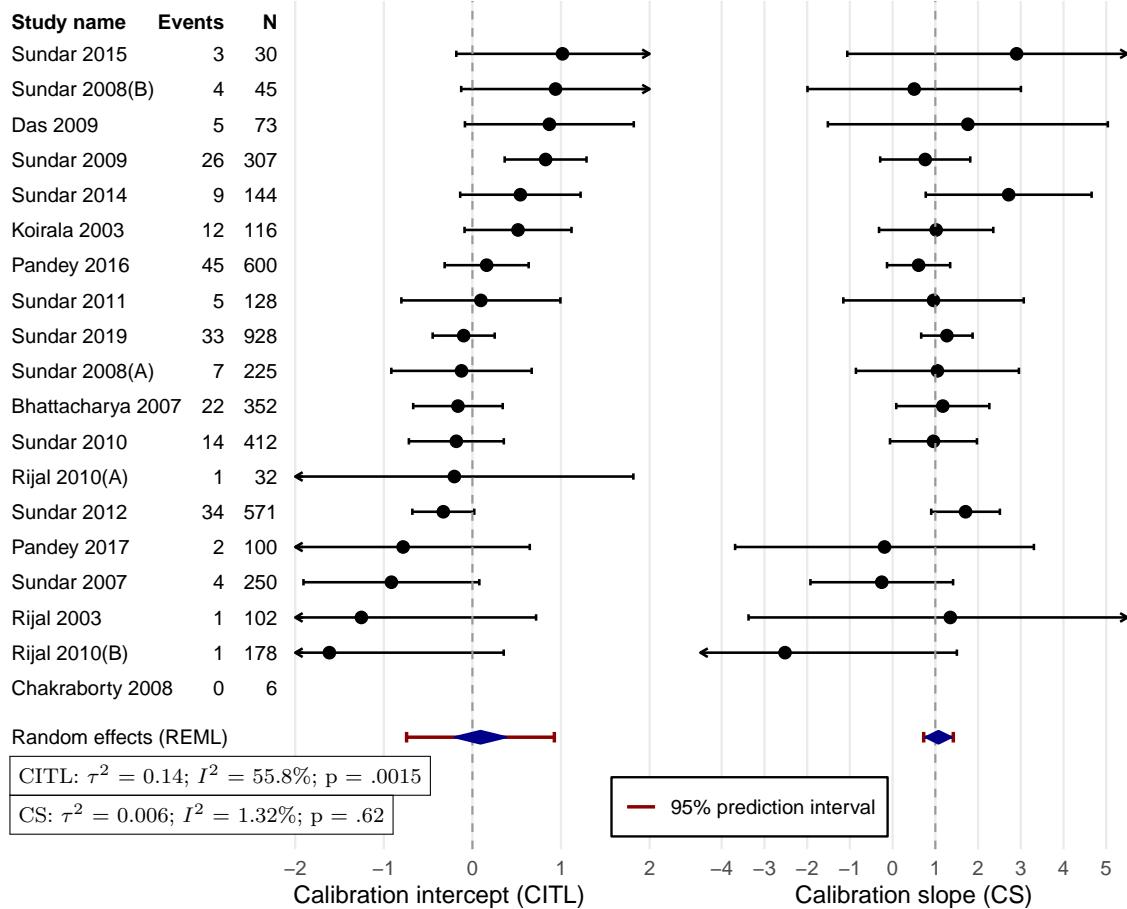


Figure C.19: Forest plots showing individual and pooled study calibration measures for the model **including** parasite grade. Left: calibration intercept (calibration-in-the-large, CITL); Right: calibration slope (CS). Blue diamonds: summary estimates with 95% confidence intervals. Calibration measures not presented for Chakraborty 2008 due to no relapse events. Calibration slope not presented for Rijal 2010(A) due to only one relapse event and few total participants leading to failure of model convergence.

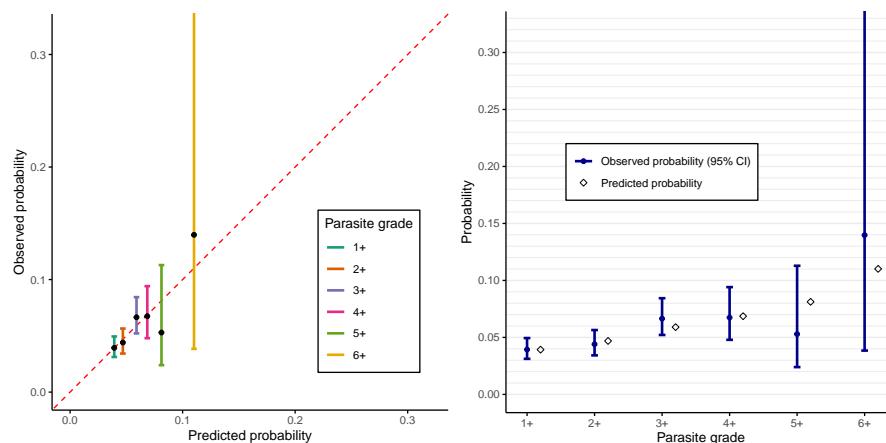


Figure C.20: Calibration plots for parasite grades — model **including** parasite grade.

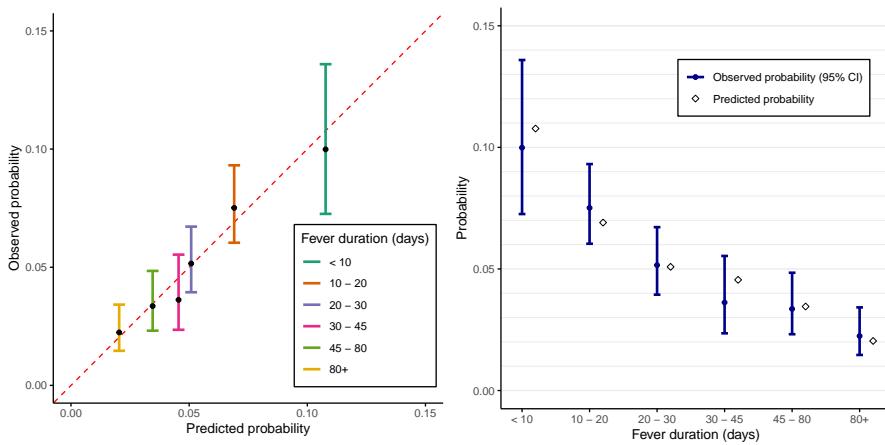


Figure C.21: Calibration plots for fever duration — model **including** parasite grade.

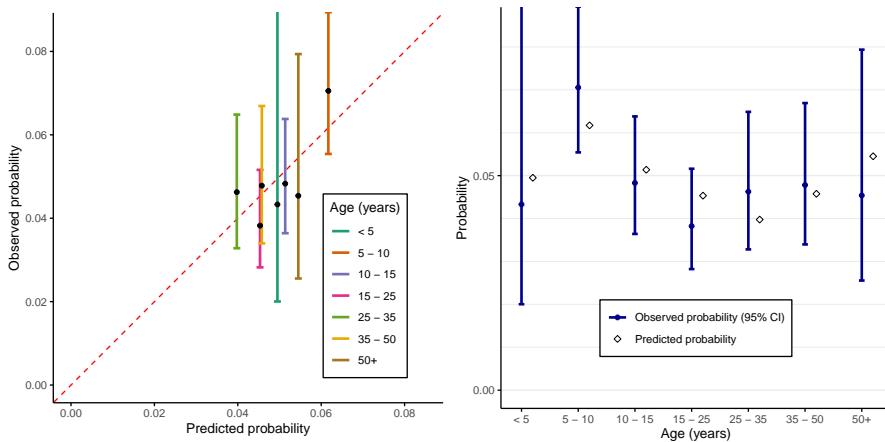


Figure C.22: Calibration plots for age — model **including** parasite grade.

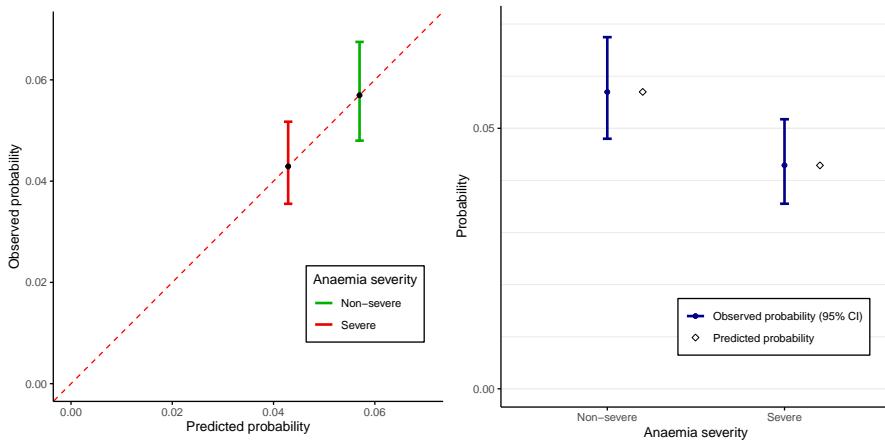


Figure C.23: Calibration plots for anaemia severity — model **including** parasite grade.

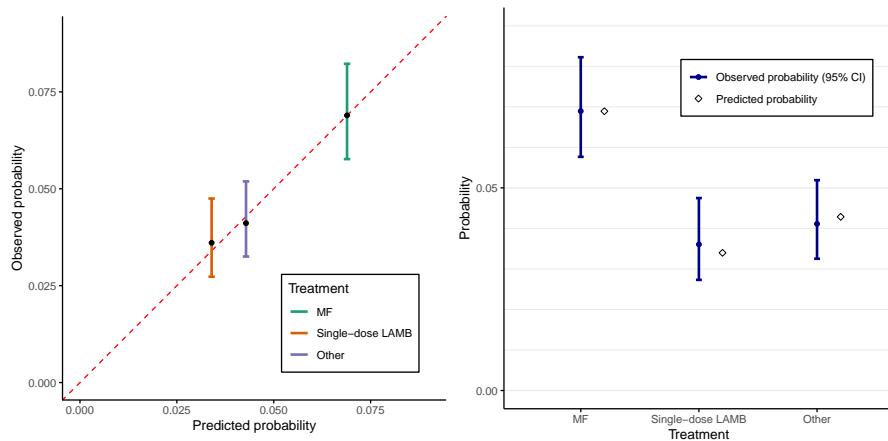


Figure C.24: Calibration plots for treatment — model **including** parasite grade.

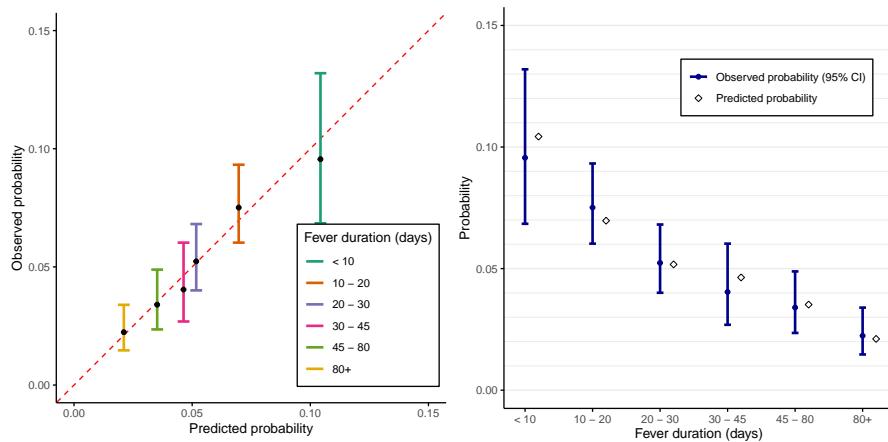


Figure C.25: Calibration plots for fever duration — model **excluding** parasite grade.

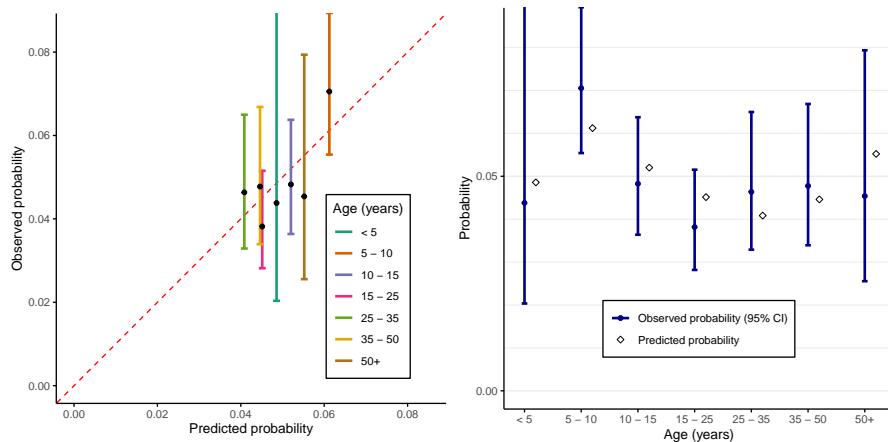


Figure C.26: Calibration plots for age — model **excluding** parasite grade.

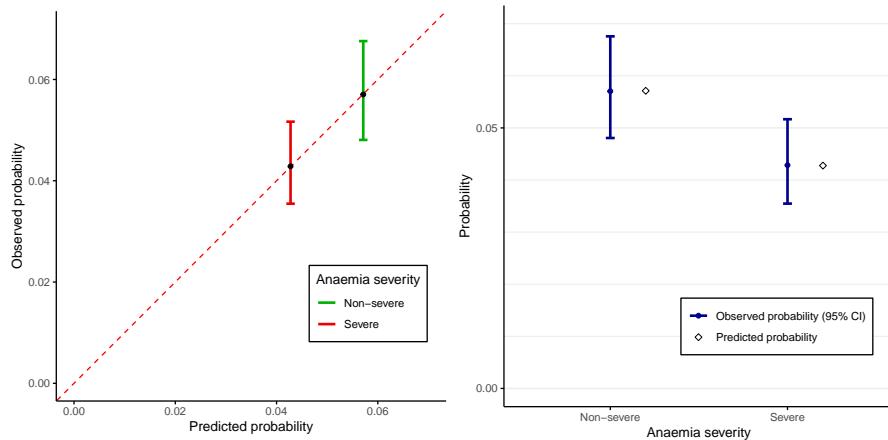


Figure C.27: Calibration plots for anaemia severity — model **excluding** parasite grade.

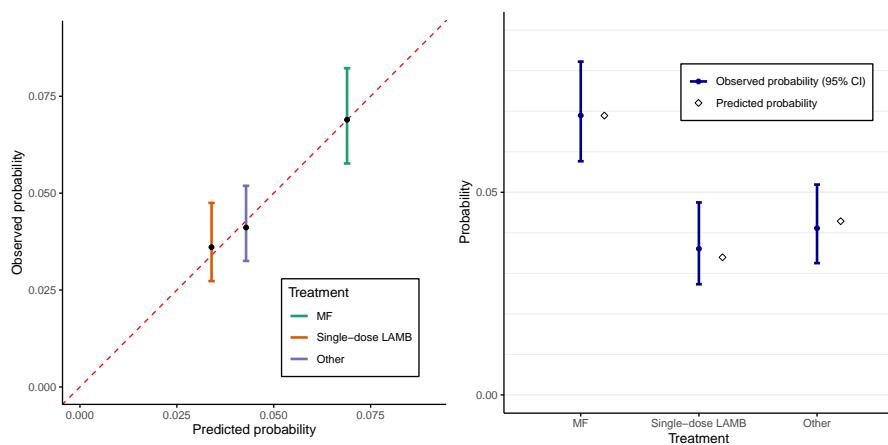


Figure C.28: Calibration plots for treatment — model **excluding** parasite grade.

D

Appendix — East Africa model results

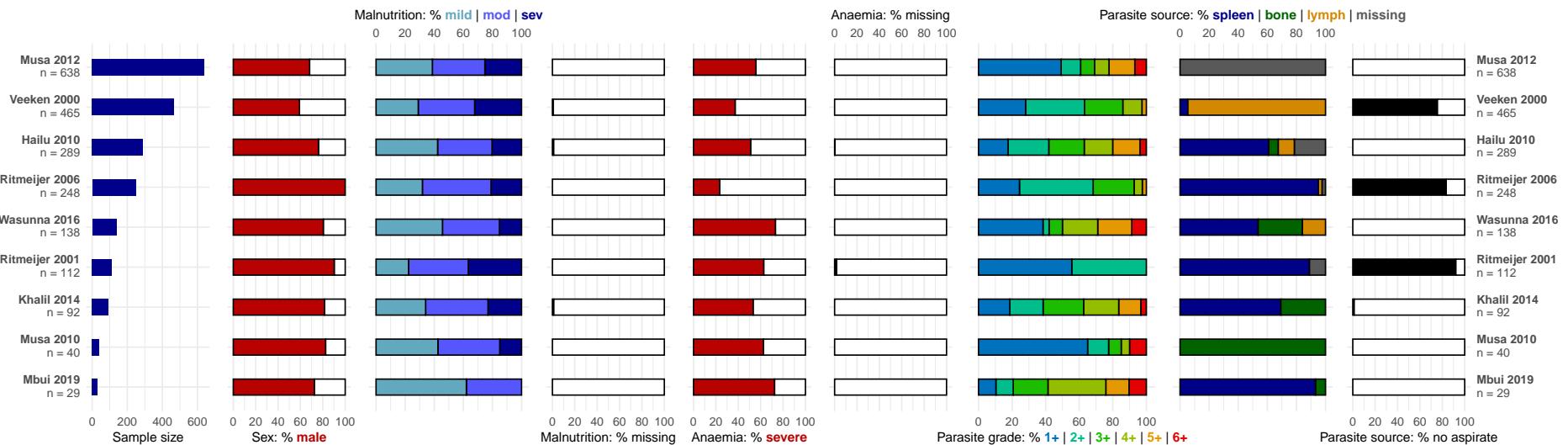


Figure D.1: Distribution of categorical predictors by study. Missing data excluded from stacked bar charts, except for parasite source, where ‘missing’ refers to patients where aspirates were performed, but aspirate source not available.

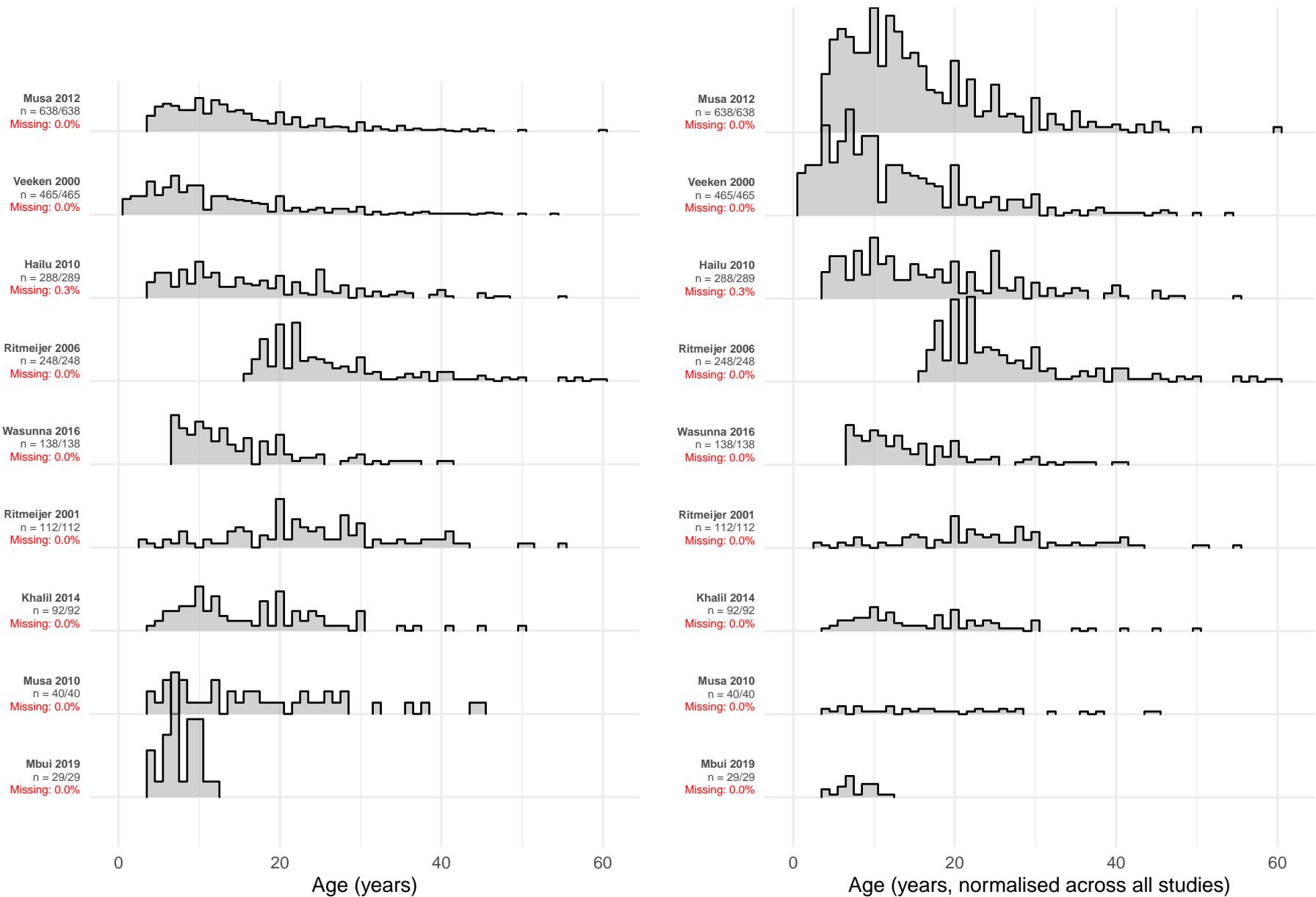


Figure D.2: Distribution of age across studies from East Africa. Missing age data described by study.

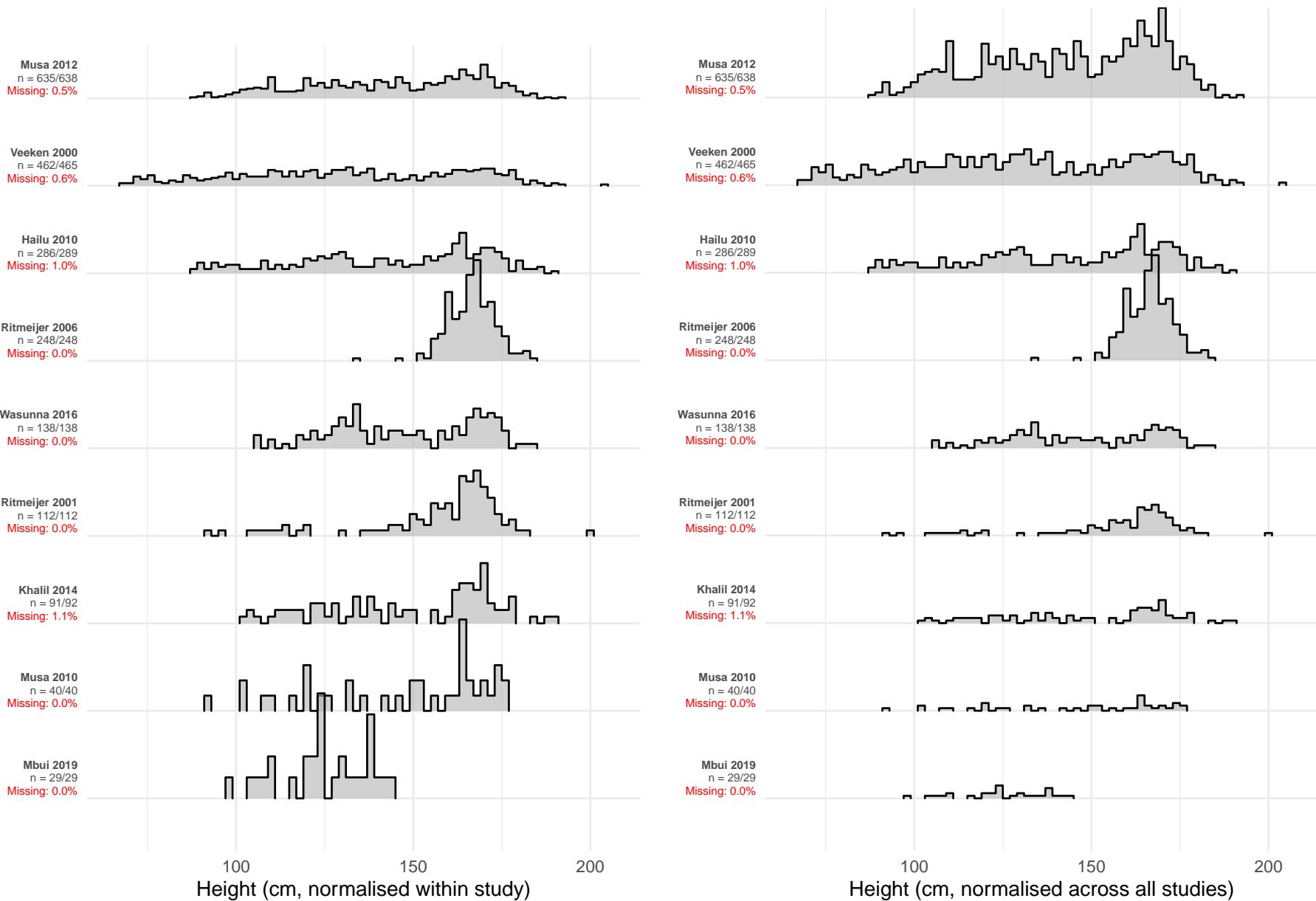


Figure D.3: Distribution of height across studies from East Africa. Missing data described by study.

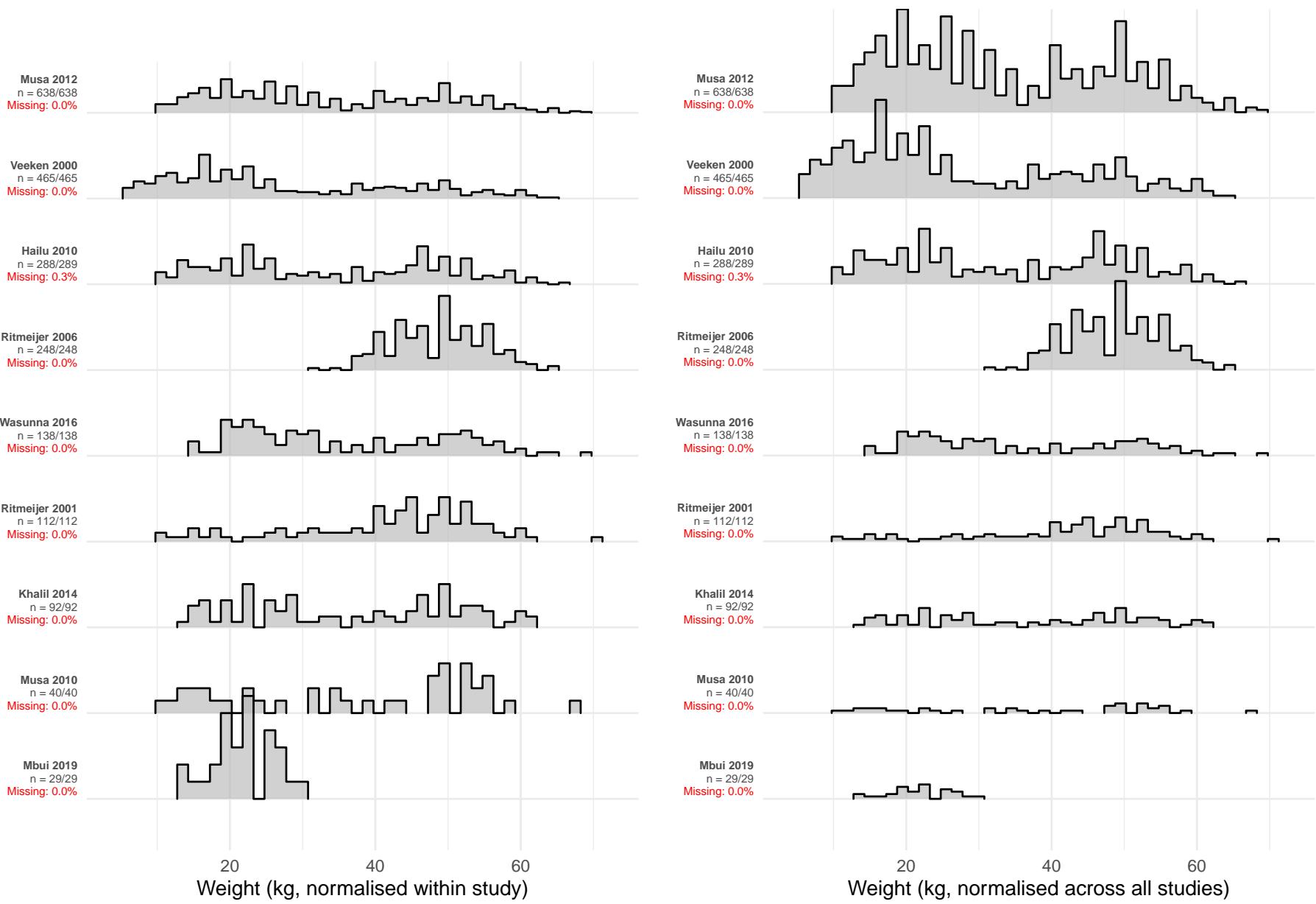


Figure D.4: Distribution of weight across studies from East Africa. Missing data described by study.

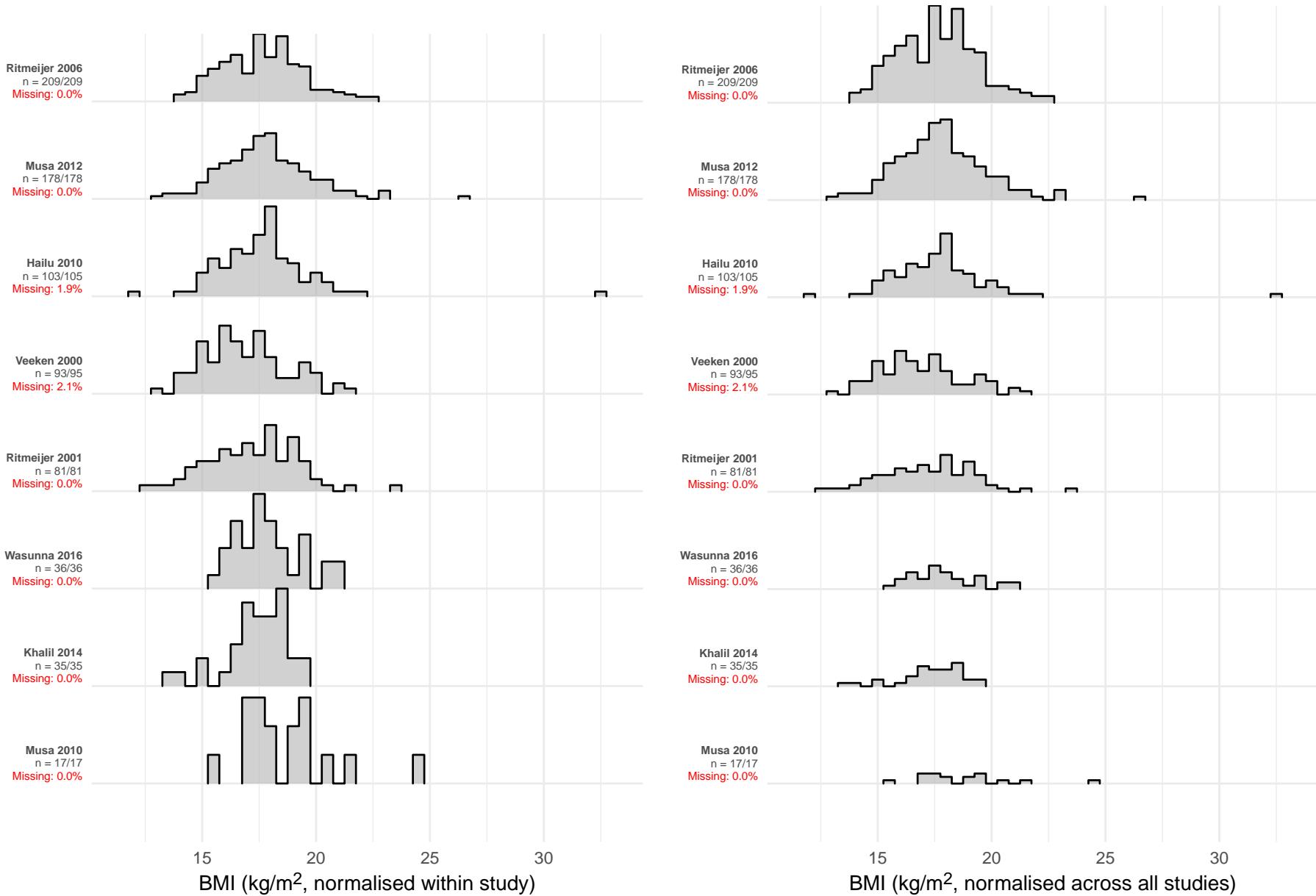


Figure D.5: Distribution of BMI across studies from East Africa. Including only participants aged 19 and over (Pandey 2017 excluded). Missing data described by study.

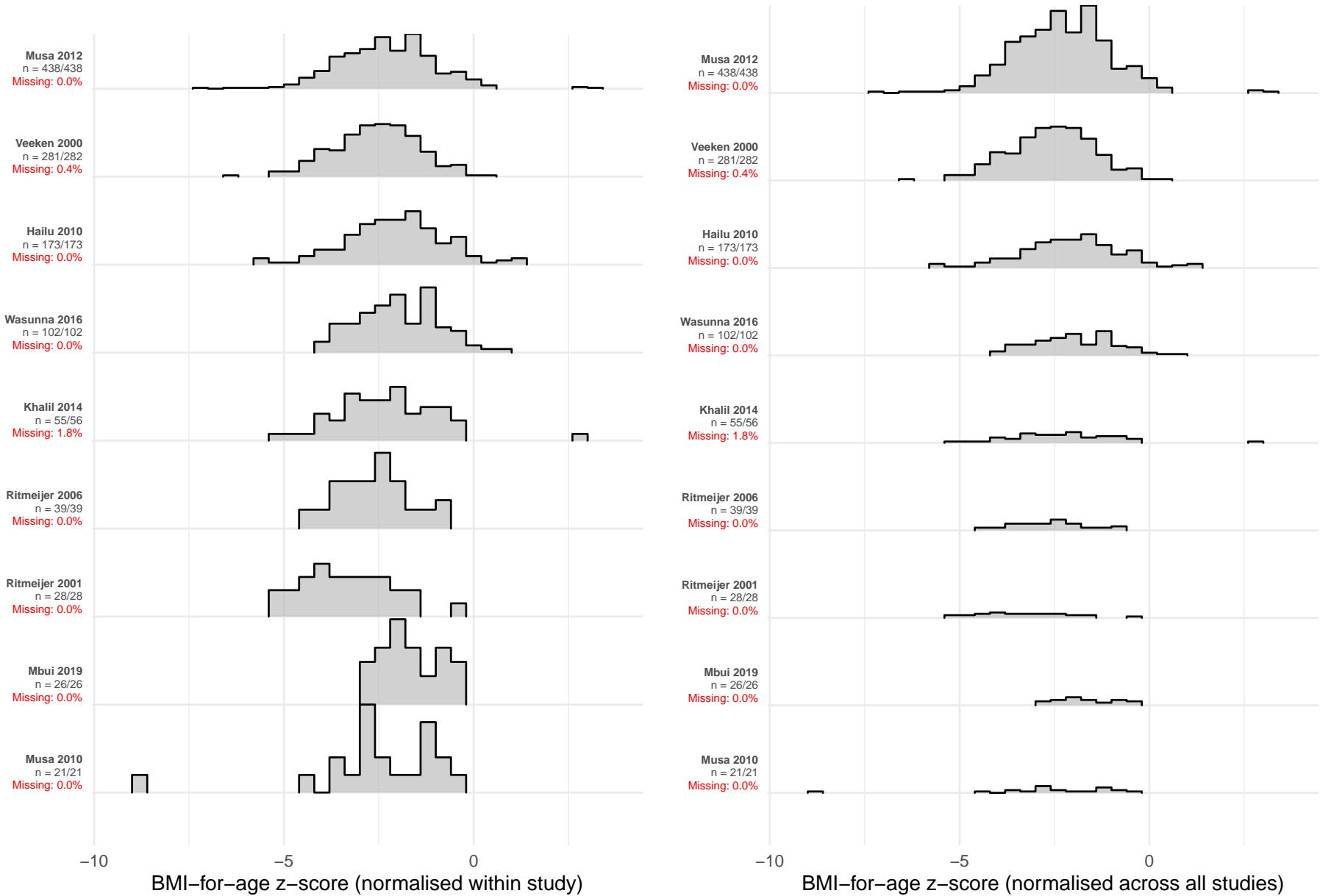


Figure D.6: Distribution of BMI-for-age z score across studies from East Africa. Including only participants aged from 5–18, inclusive. Missing data described by study.

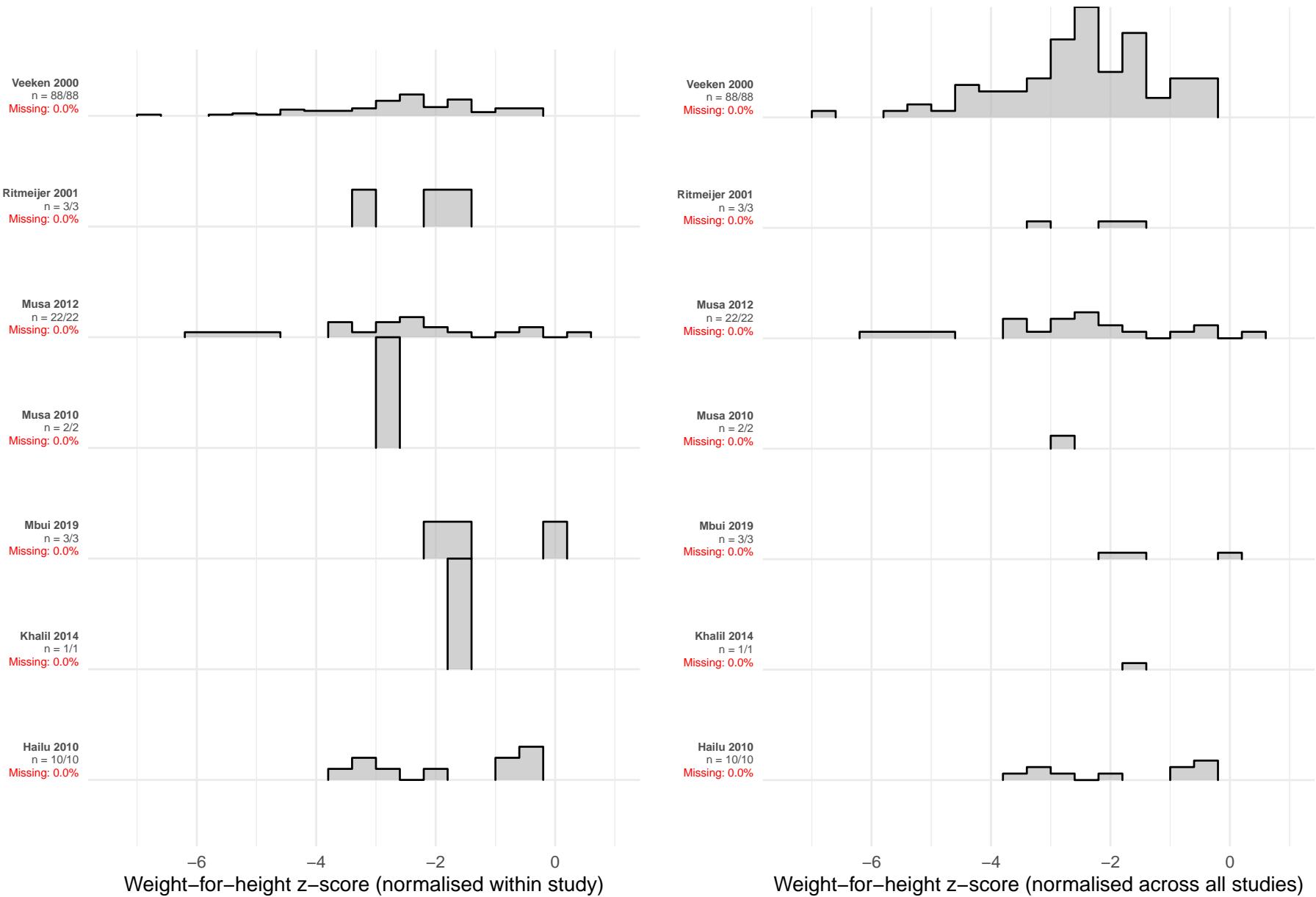


Figure D.7: Distribution of weight-for-height z score across studies from East Africa. Including only participants aged under 5. Missing data described by study.

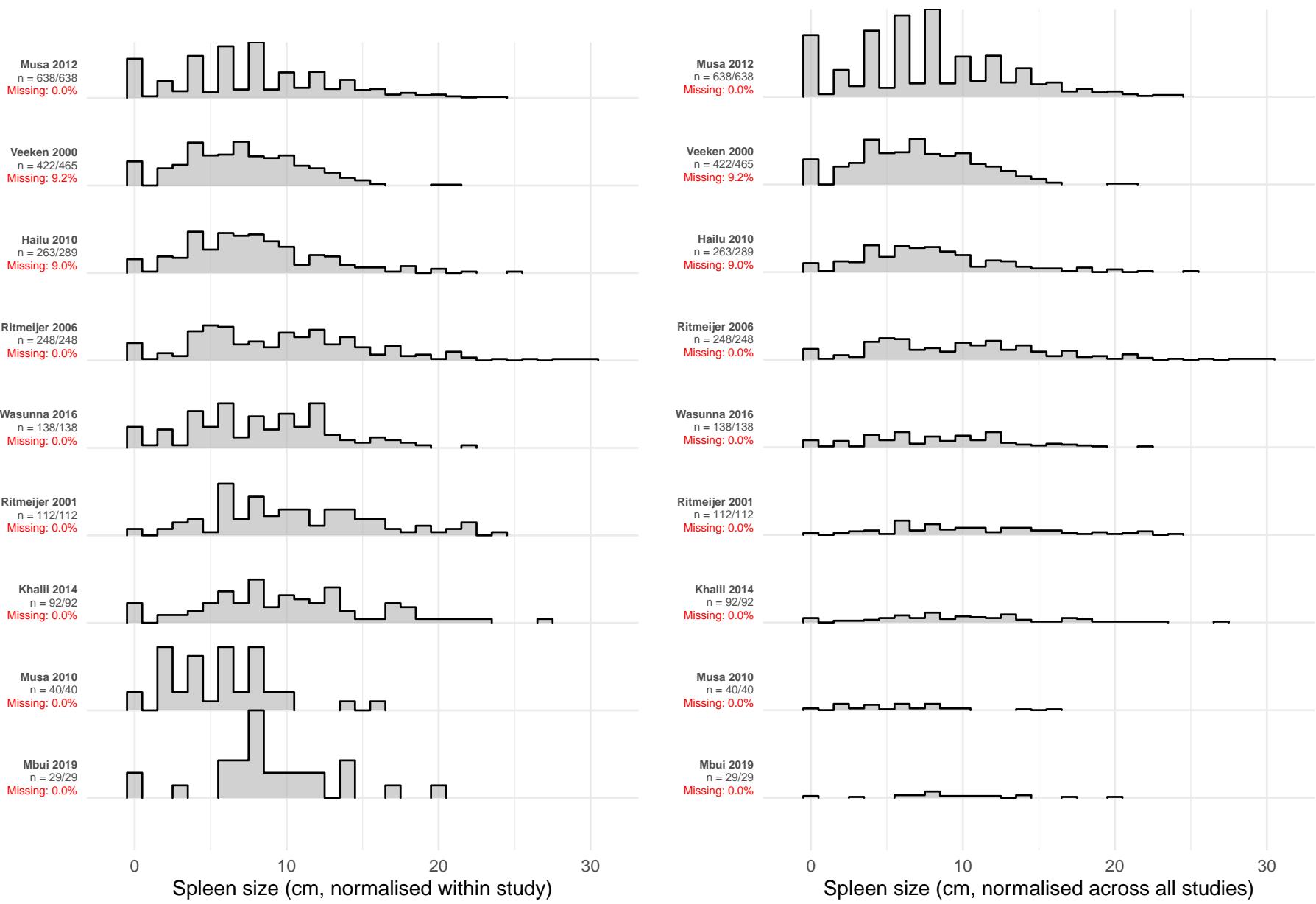


Figure D.8: Distribution of spleen size across studies from East Africa. Missing data described by study.

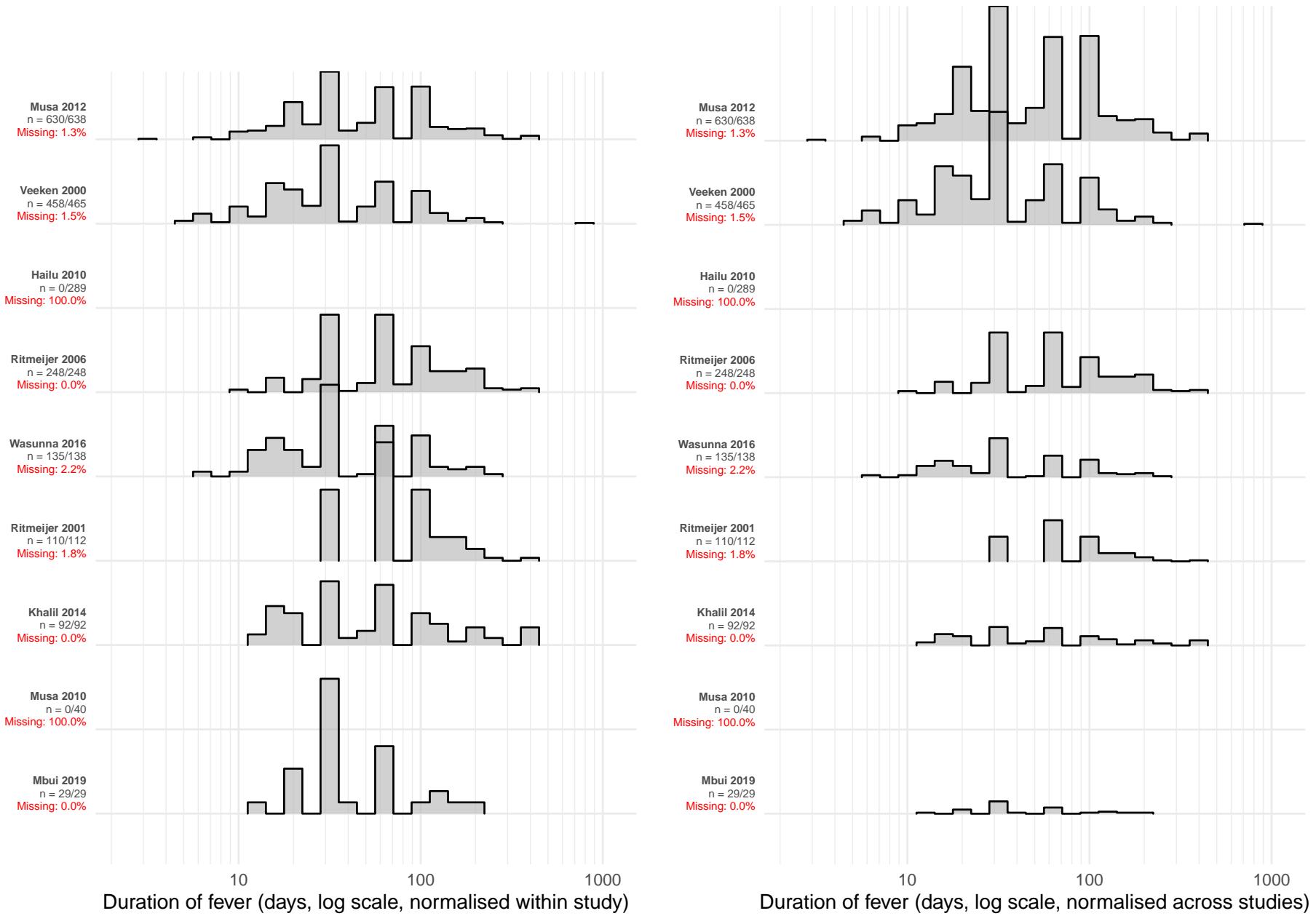


Figure D.9: Distribution of fever duration (log scale) across studies from East Africa. Missing data described by study.

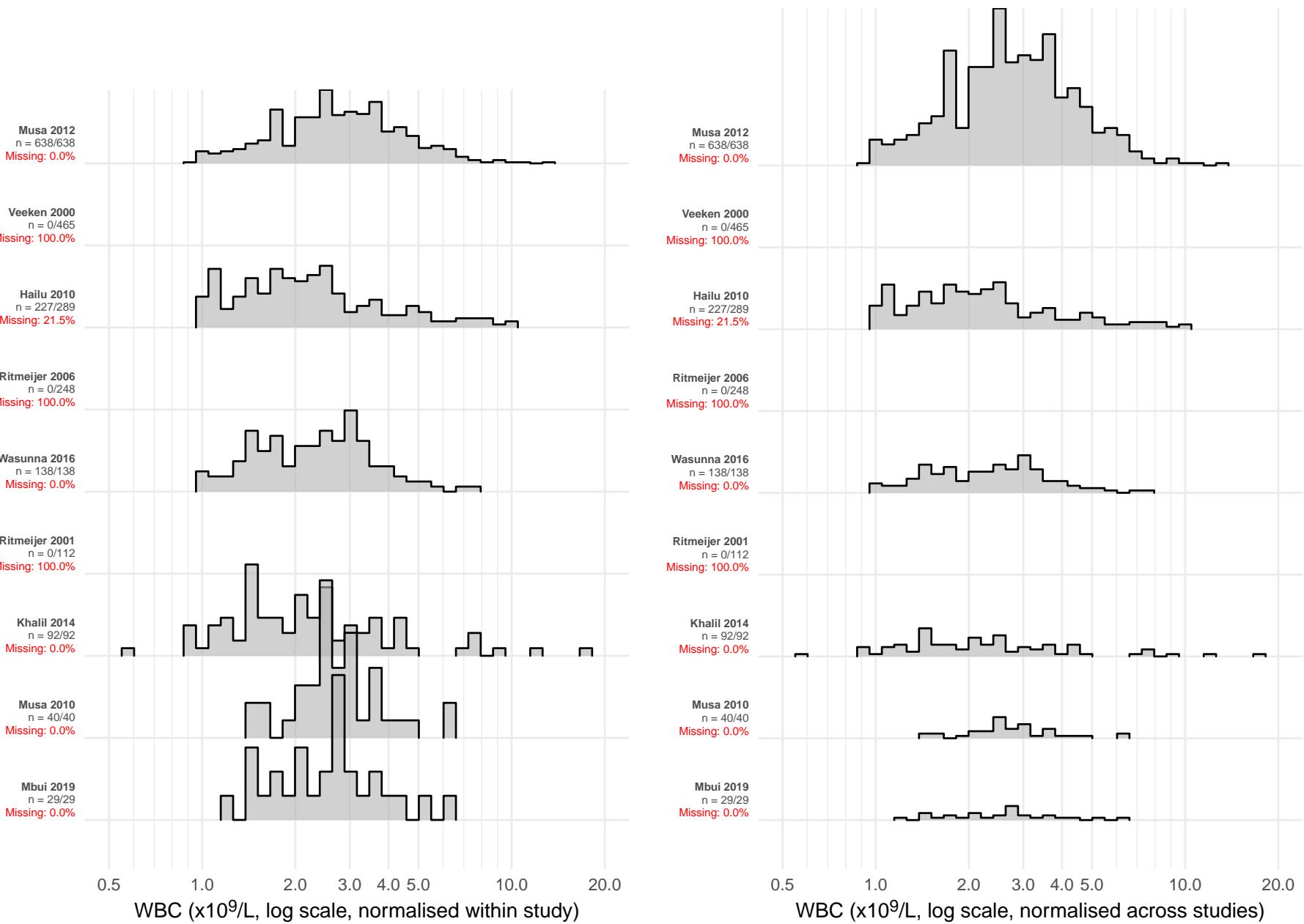


Figure D.10: Distribution of white blood cell count (log scale) across studies from East Africa. Missing data described by study.

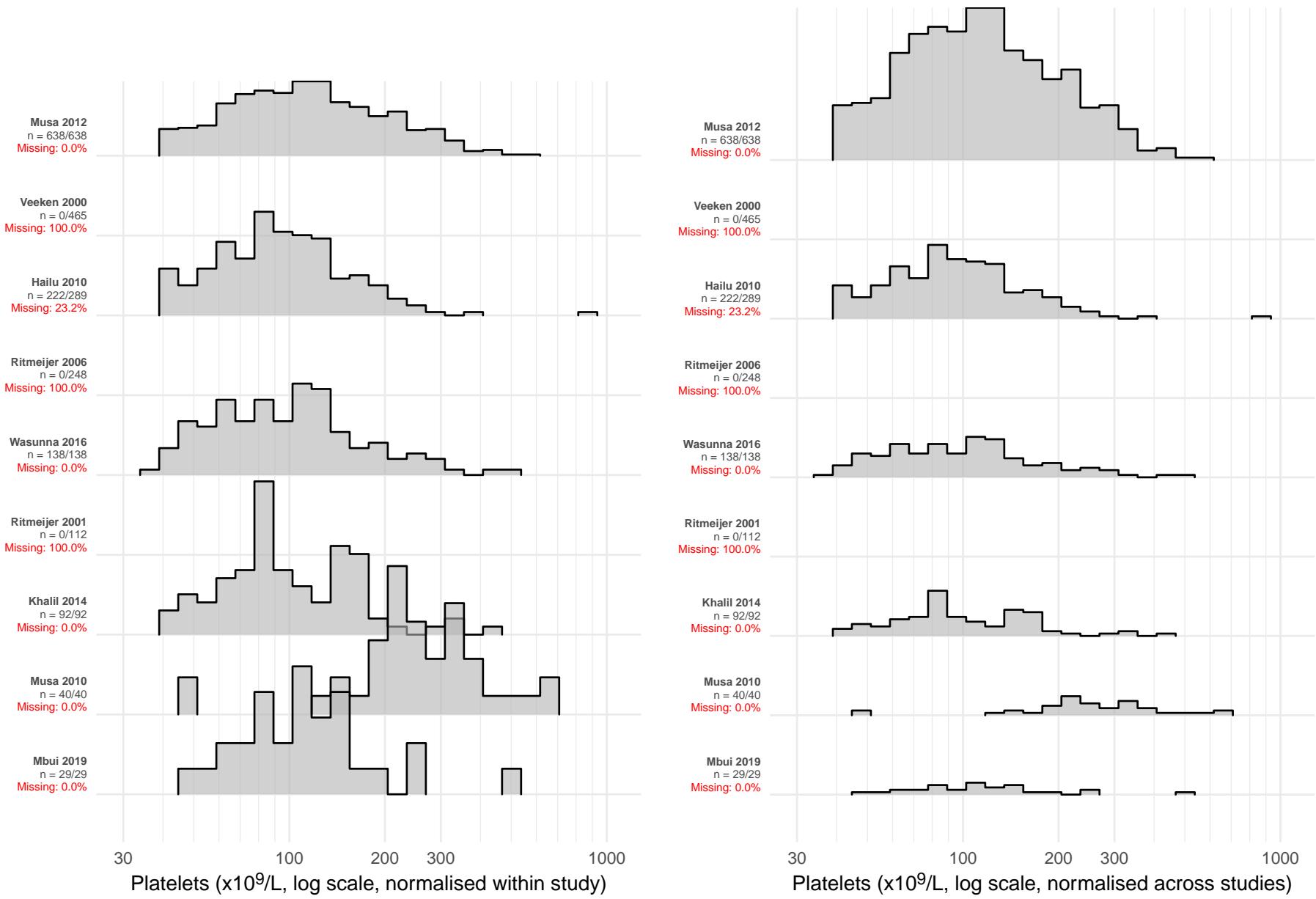


Figure D.11: Distribution of platelet count (log scale) across studies from East Africa. Missing data described by study.

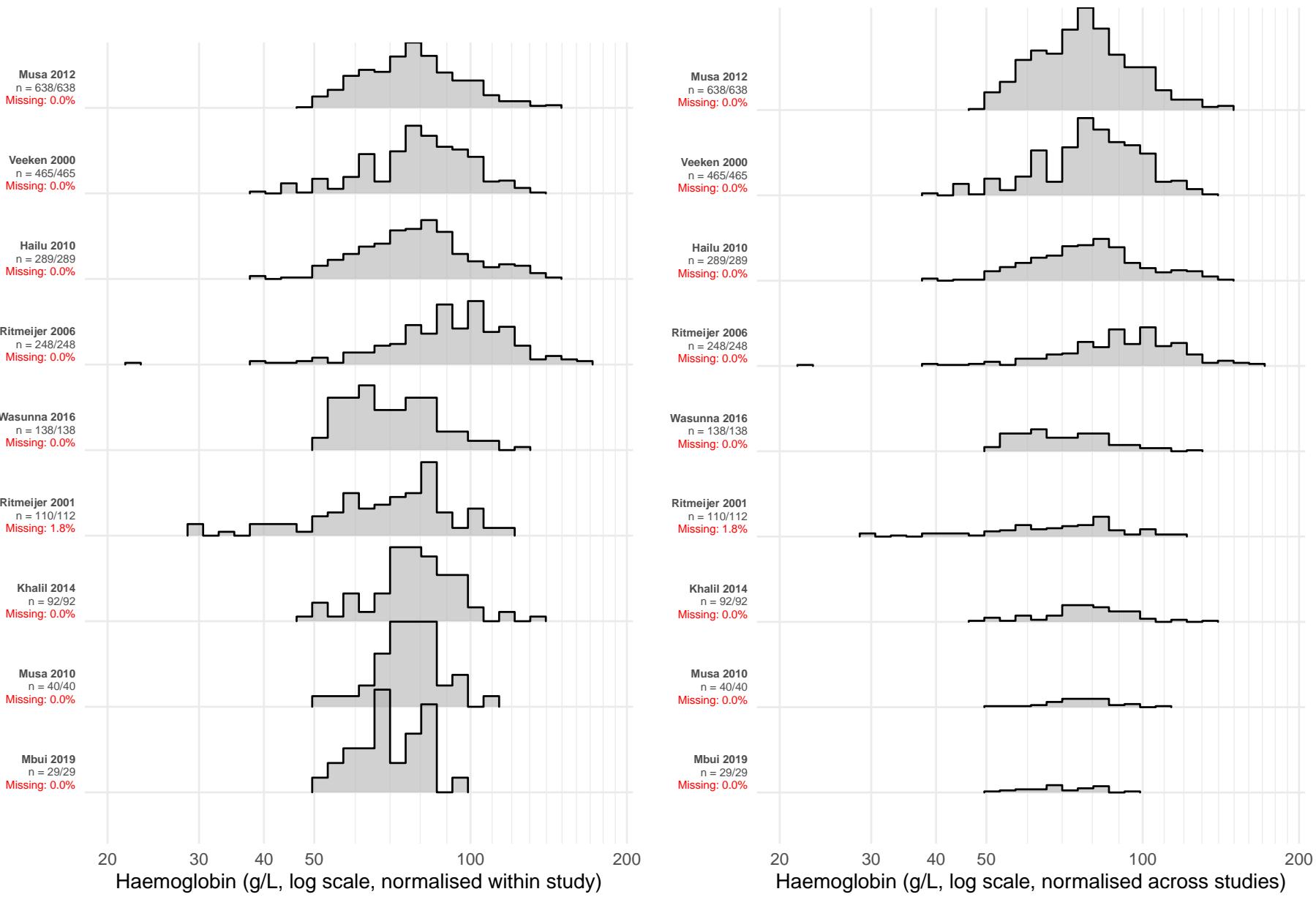


Figure D.12: Distribution of haemoglobin (log scale) across studies from East Africa. Missing data described by study.

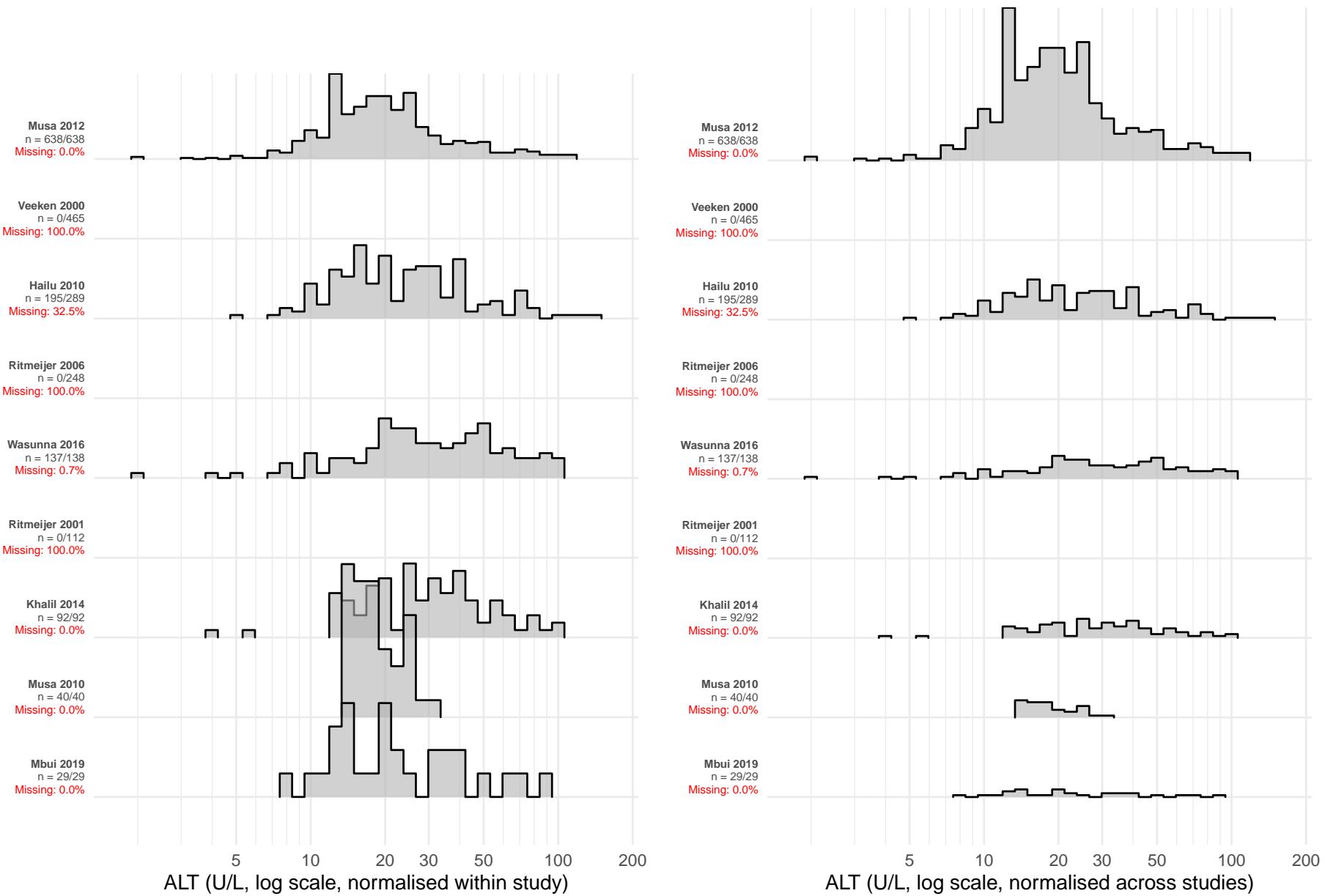


Figure D.13: Distribution of alanine transaminase (ALT, log scale) across studies from East Africa. Missing data described by study.

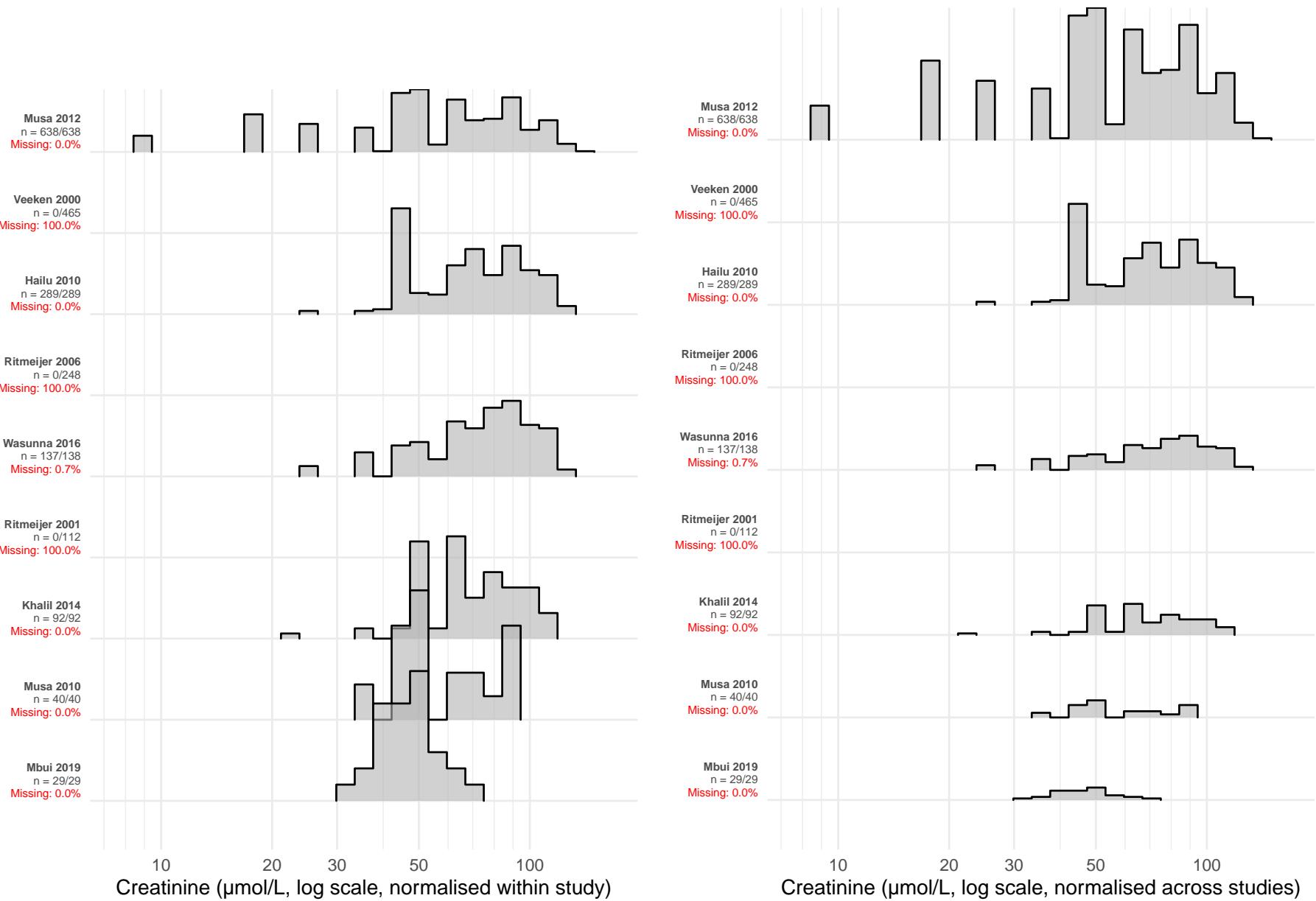


Figure D.14: Distribution of creatinine (log scale) across studies from East Africa. Missing data described by study.

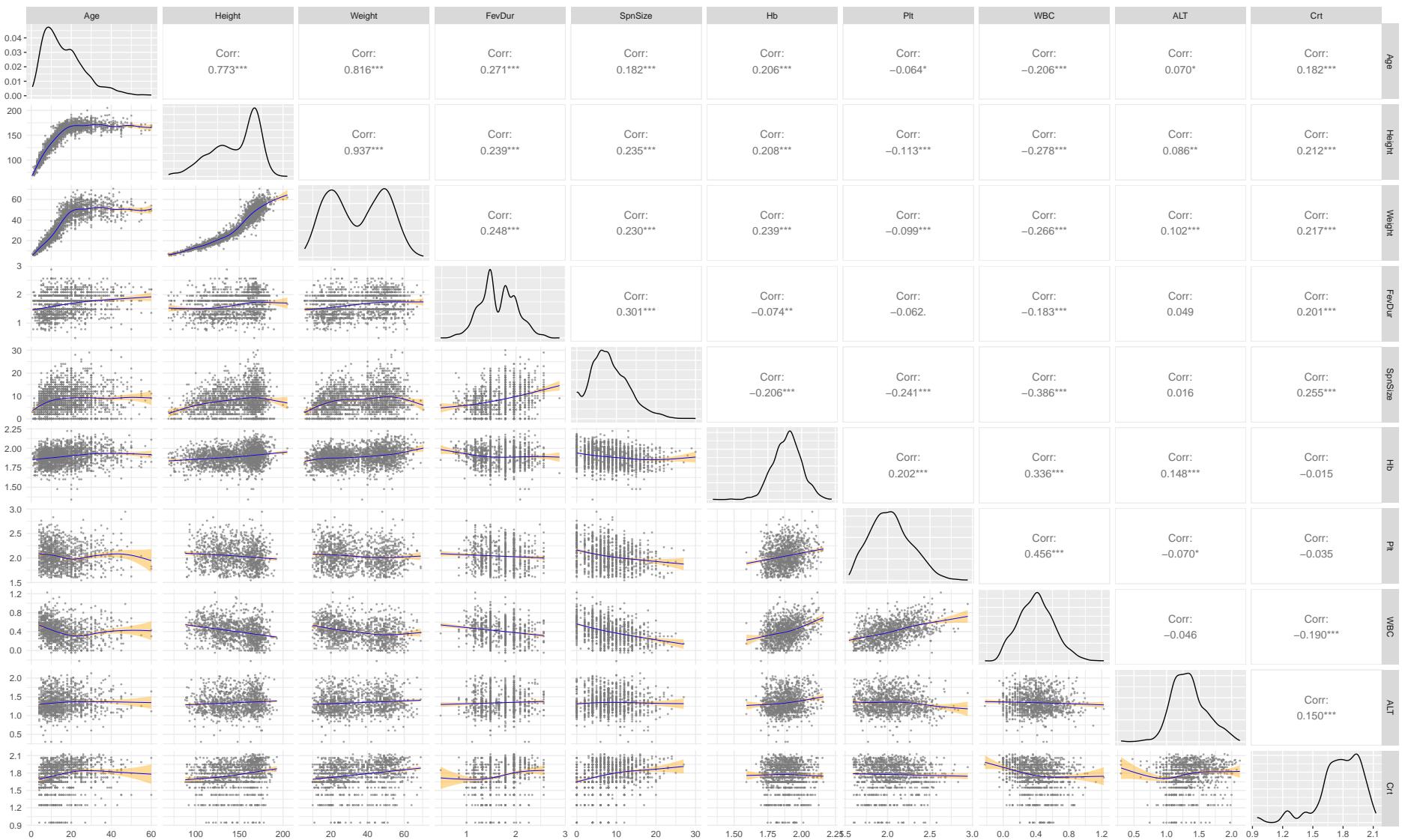


Figure D.15: Correlation between continuous variables. For scatter plots, a univariable generalised additive model is fitted (blue line) with 95% confidence interval ribbon filled (orange area). Pearson correlation coefficients are presented, *** p < 0.001; ** p < 0.01; * p < 0.05, ' p < 0.10. Age: years; Height: cm; Weight: kg; FevDur: duration of fever, log10(days); SpnSize: spleen size (cm); Hb: haemoglobin, log10(g/L); Plt: platelets, log10($\times 10^9/L$); WBC: white blood cells log10($\times 10^9/L$); ALT: alanine aminotransferase, log10(U/L); Cr: creatinine log10($\mu\text{mol}/L$).

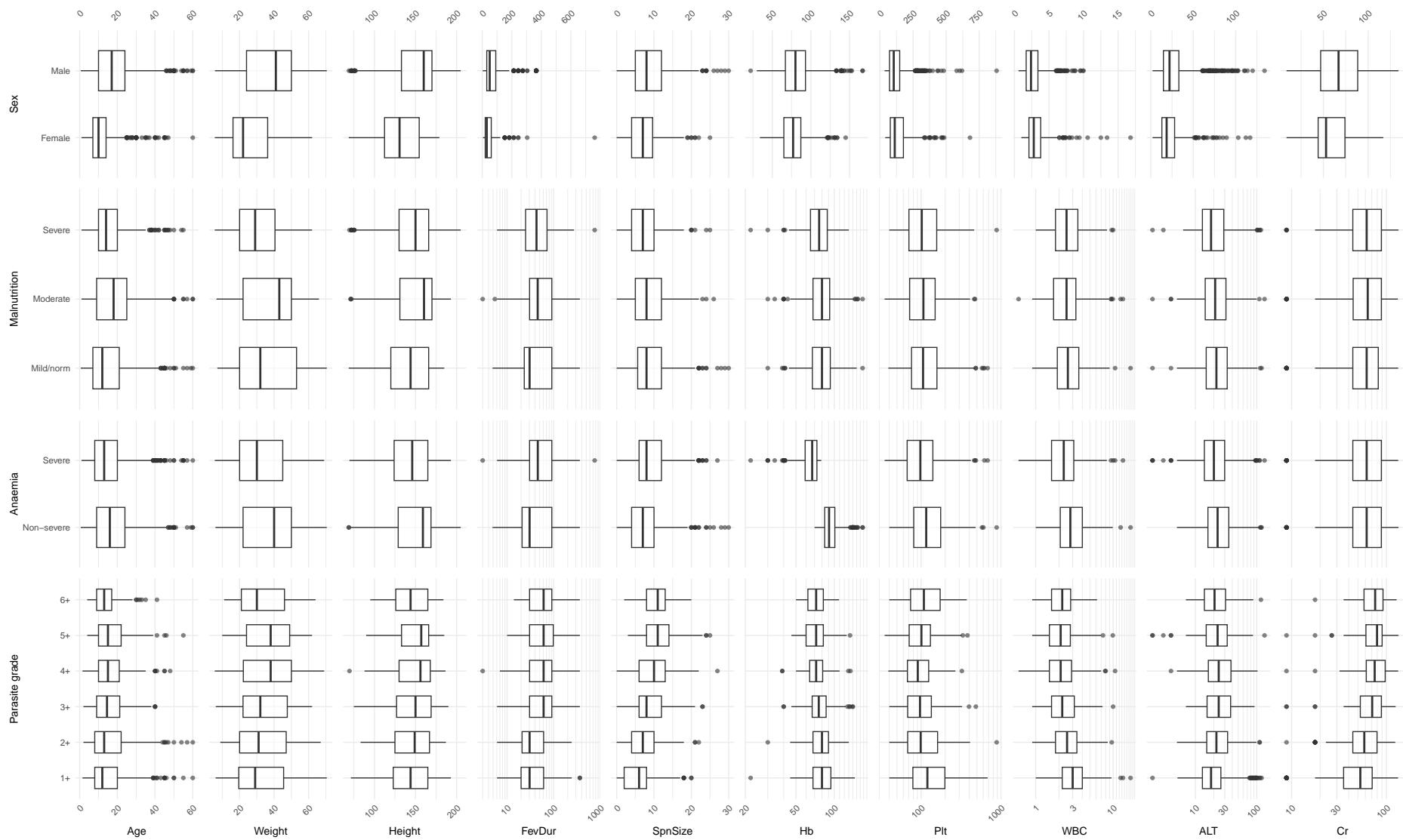


Figure D.16: Correlation between continuous and categorical variables. FevDur and laboratory tests axes transformed to log₁₀ scale. Age: years; Height: cm; Weight: kg; FevDur: duration of fever, days; SpnSize: spleen size, cm; Hb: haemoglobin, g/L; Plt: platelets, $\times 10^9/L$; WBC: white blood cells $\times 10^9/L$; ALT: alanine aminotransferase, U/L; Cr: creatinine $\mu\text{mol}/L$.

term	estimate ¹	std.error	t statistic	df (effective)	p value
Fixed intercept	-3.7874	0.5369	-7.05	1885.50	2.43E-12
Parasite grade ²	0.3593	0.0763	4.71	1044.63	2.83E-06
Spleen size ³	-0.3335	0.1741	-1.92	1813.86	5.56E-02
Anaemia: Severe ⁴	0.5450	0.2448	2.23	1973.14	2.61E-02
Fever duration ⁵	-0.3302	0.1349	-2.45	476.32	1.47E-02
WBC count ⁶	0.2368	0.1315	1.80	419.77	7.23E-02
Malnutrition: Moderate ⁷	0.5769	0.2677	2.16	2004.45	3.13E-02
Malnutrition: Severe ⁷	0.9780	0.2860	3.42	2025.03	6.39E-04
Unadjusted intercept (Musa 2012):	-3.9411				
Uniform shrinkage factor:	0.8079				
Adjusted intercepts after shrinkage: Average:	-4.3182; Musa 2012:	-4.4524			
Pooled τ^2 :	0.9122.	Pooled ICC:	0.2169		

¹ Logit scale before uniform shrinkage.

² Continuous variable, from 1+ to 6+ as described by Chulay and Bryceson[52]. Transformed $x \mapsto x - 1$.

³ Units: centimetres. Converted to natural log scale $x \mapsto \log(x + 1)$.

⁴ Reference group: non-severe anaemia.

⁵ Units: days. Converted to natural log scale and then standardised by mean (3.550509) and standard deviation (0.8693318).

⁶ Units: $\times 10^9$ cells per litre. Converted to natural log scale and then standardised by mean (0.926133) and standard deviation (0.4911651).

⁷ Reference group: Mild/normal malnutrition.

Table D.1: Model coefficients pooled across 30 imputed datasets using Rubin's Rules.

Model **including** parasite grade. df: degrees of freedom; std.error: standard error; WBC: white blood cell count

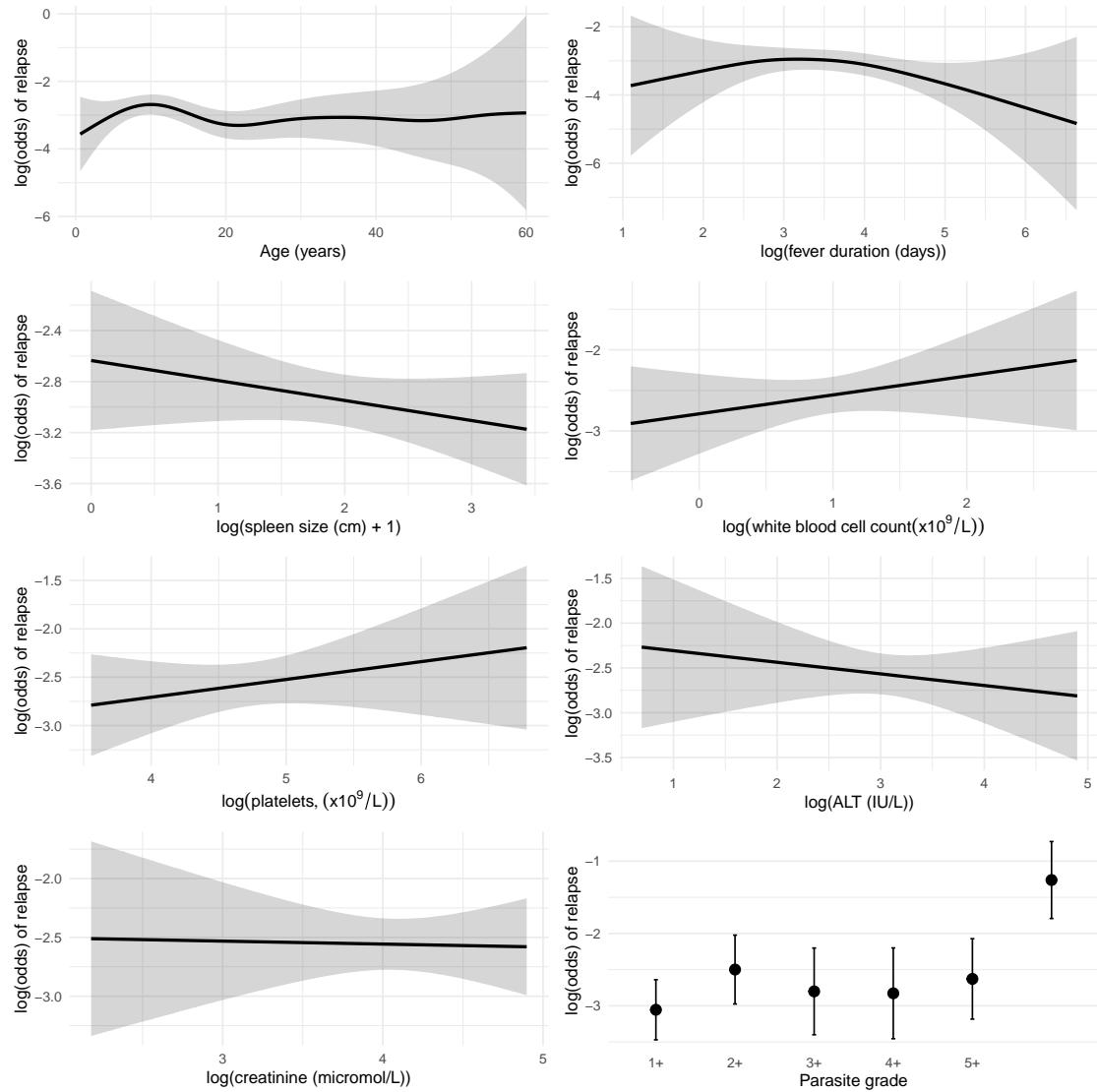


Figure D.17: Associations between transformed continuous predictors and $\log(\text{odds})$ of relapse. For each predictor (excluding parasite grade), a univariable generalised additive model spline fit is shown, with 95% confidence intervals. For parasite grade, 95% confidence intervals are calculated for each grade using the Wilson method.

term	estimate ¹	std.error	t statistic	df (effective)	p value
Fixed intercept	-3.7674	0.4381	-8.60	2028.60	1.57E-17
Anaemia: Severe ²	0.5553	0.2365	2.35	1981.61	1.89E-02
WBC count ³	0.2254	0.1167	1.93	715.01	5.39E-02
Malnutrition: Moderate ⁴	0.5486	0.2625	2.09	2002.50	3.67E-02
Malnutrition: Severe ⁴	0.9282	0.2794	3.32	2017.72	9.08E-04
Unadjusted intercept (Musa 2012): -3.9073					
Uniform shrinkage factor: 0.7493					
Adjusted intercepts after shrinkage: Average: -4.4889; Musa 2012: -4.6065					
Pooled τ^2 : 1.040. Pooled ICC: 0.2401					

¹ Logit scale before uniform shrinkage.

² Reference group: non-severe anaemia.

³ Units: $\times 10^9$ cells per litre. Converted to natural log scale and then standardised by mean (0.926133) and standard deviation (0.4911651).

⁴ Reference group: Mild/normal malnutrition.

Table D.2: Model coefficients pooled across 30 imputed datasets using Rubin's Rules. Model **excluding** parasite grade. df: degrees of freedom; std.error: standard error; WBC: white blood cell count

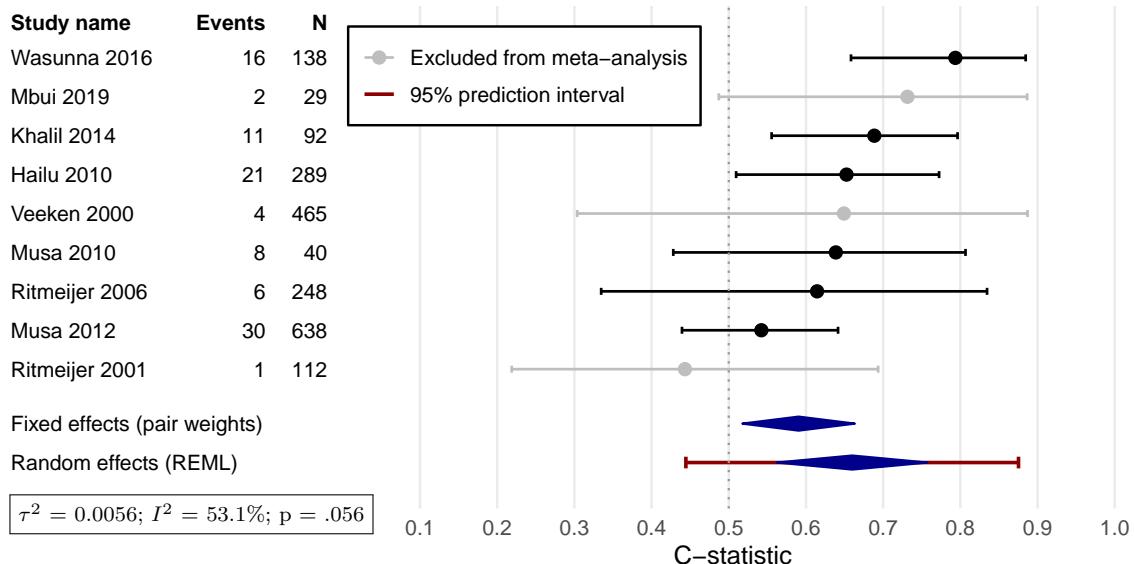


Figure D.18: Forest plot showing individual and pooled study c-statistics, for the model **excluding** parasite grade. Pooled c-statistics are presented from both fixed-effects and random-effects meta-analysis models. Pooled random-effects c-statistics and variances are estimated using restricted maximum likelihood (REML) and the Hartung-Knapp-Sidik-Jonkman method. Study-specific confidence intervals should be interpreted with caution due to small sample sizes and relapse events in some studies (see Methodology Section 4.3.7). Blue diamonds: pooled summary estimates with 95% confidence intervals. Red line: 95% prediction interval. Study ordered by c-statistic.

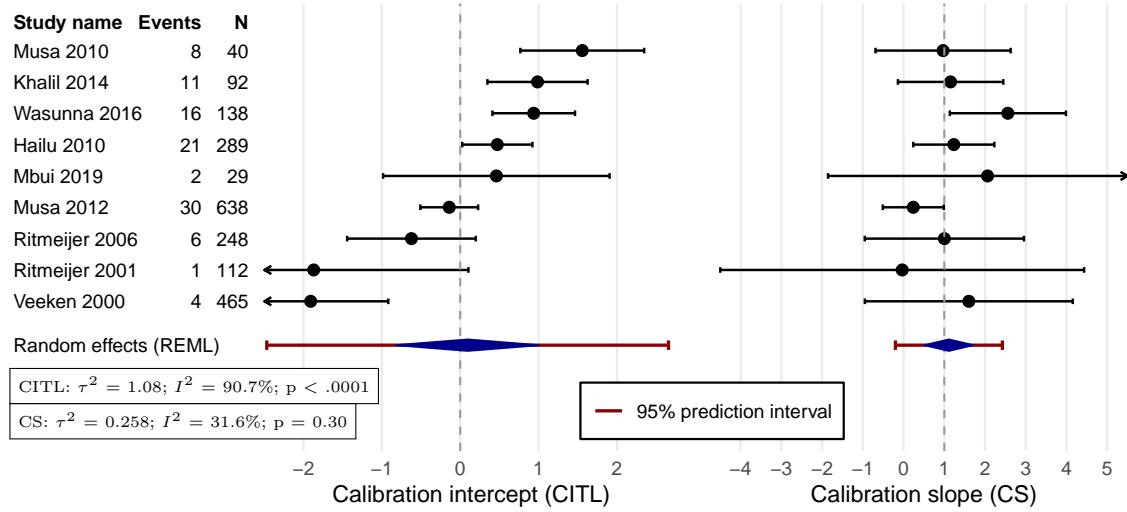


Figure D.19: Forest plots showing individual and pooled study calibration measures for the model **excluding** parasite grade. Left: calibration intercept (calibration-in-the-large, CITL); Right: calibration slope (CS). Blue diamonds: summary estimates with 95% confidence intervals.

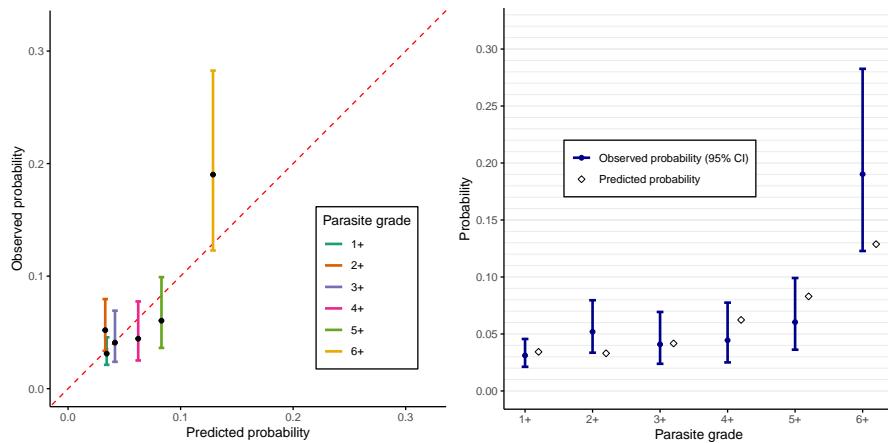


Figure D.20: Calibration plots for parasite grades — model **including** parasite grade.

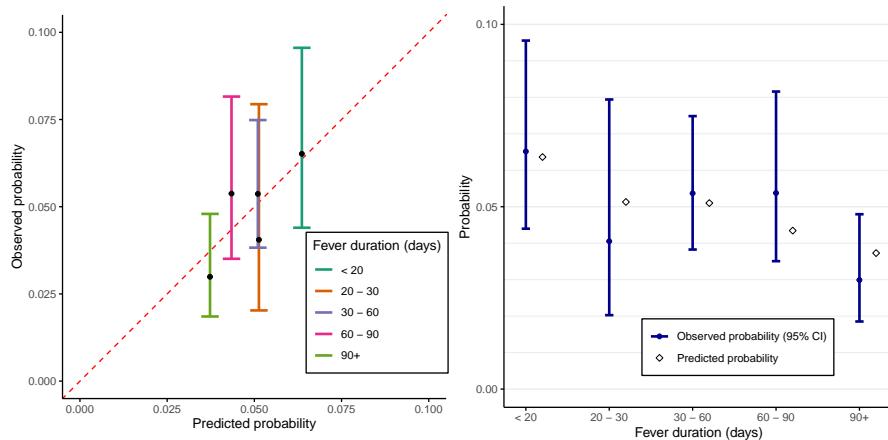


Figure D.21: Calibration plots for fever duration — model **including** parasite grade.

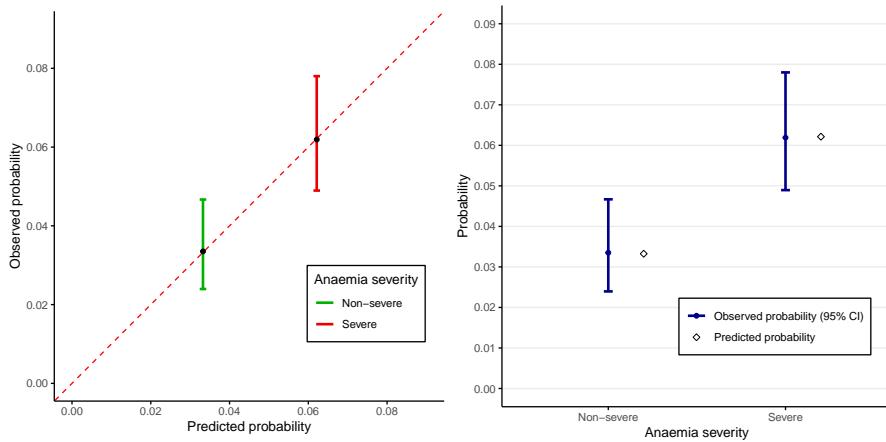


Figure D.22: Calibration plots for anaemia — model **including** parasite grade.

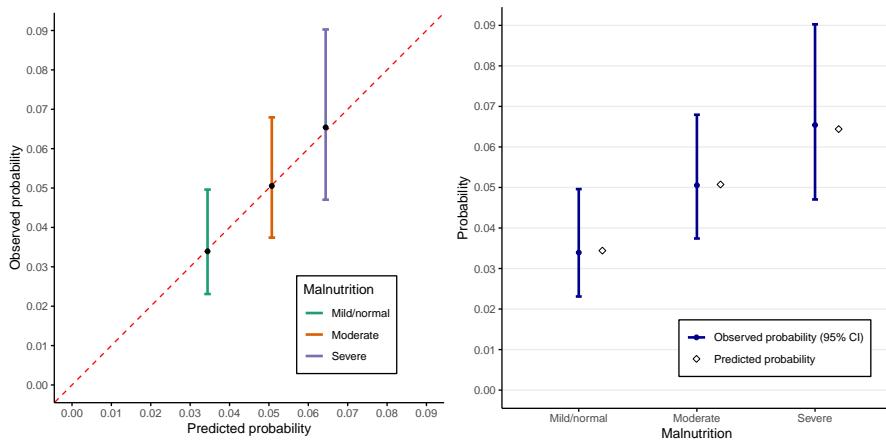


Figure D.23: Calibration plots for malnutrition — model **including** parasite grade.

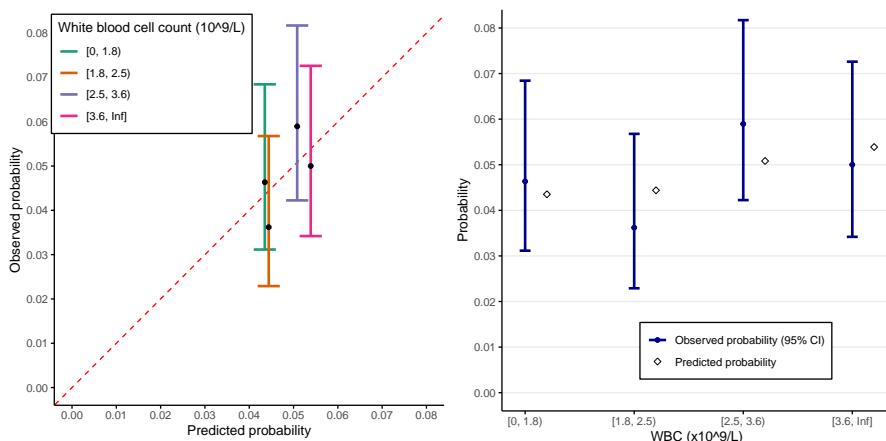


Figure D.24: Calibration plots for white blood cell count — model **including** parasite grade.

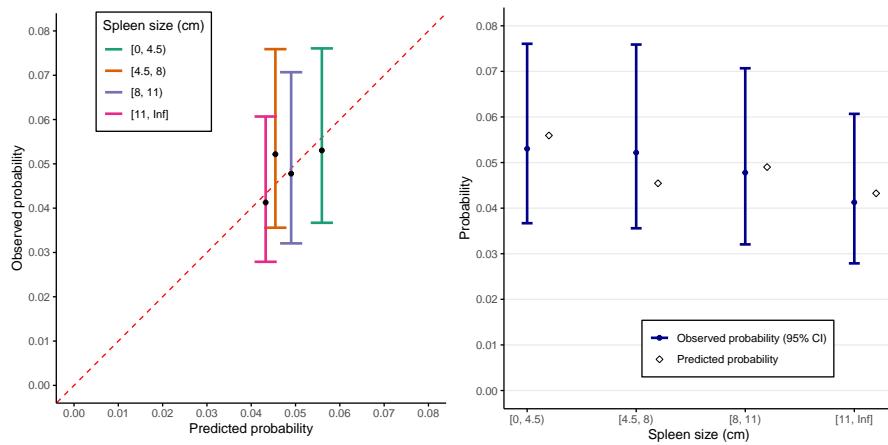


Figure D.25: Calibration plots for spleen size — model **including** parasite grade.

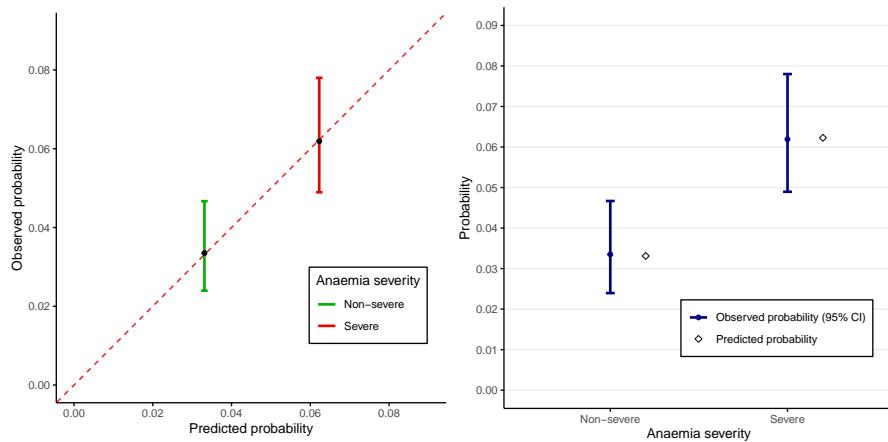


Figure D.26: Calibration plots for anaemia — model **excluding** parasite grade.

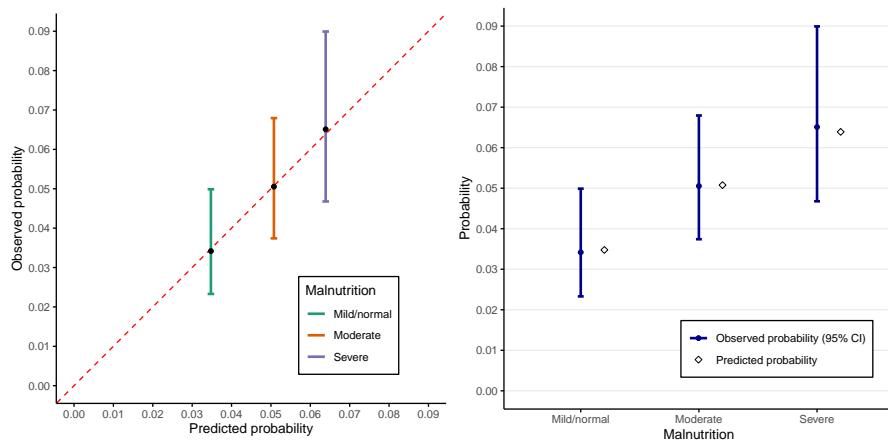


Figure D.27: Calibration plots for malnutrition — model **excluding** parasite grade.

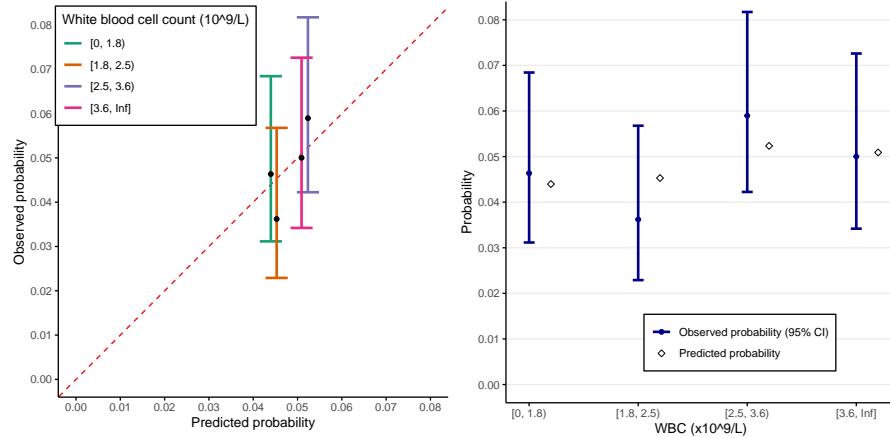


Figure D.28: Calibration plots for white blood cell count — model **excluding** parasite grade.

Predictors	Model with PG	Model without PG
Parasite grade	499¹	n/a
Malnutrition	477¹	461¹
Fever duration	390¹	176
Anaemia	298¹	361¹
Spleen size	281¹	84
WBC count	254¹	188¹
Age (linear)	205	209
Creatinine	194	101
Age (squared)	162	161
ALT	137	138
Sex	124	119
Age (cubic)	57	71
Platelets	55	53

¹ Selected in final model.

Table D.3: Frequency of predictor selection during 500 bootstrap model developments. Predictors sorted by selection frequency in the model with parasite grade. ALT: alanine aminotransferase; n/a: not applicable; PG: parasite grade; WBC: white blood cell count

References

- [1] Myrthe Pareyn et al. “Leishmaniasis”. en. In: *Nature Reviews Disease Primers* 11.1 (Nov. 2025), p. 81. DOI: [10.1038/s41572-025-00663-w](https://doi.org/10.1038/s41572-025-00663-w).
- [2] Marleen Boelaert et al. “Leishmaniasis”. In: *Manson’s Tropical Diseases: Twenty-Fourth Edition*. Ed. by Jeremy Farrar et al. London / Philadelphia: Elsevier Health Sciences, 2023. Chap. 53, pp. 646–668. ISBN: 9780702079597.
- [3] Sakib Burza et al. “Leishmaniasis”. In: *Lancet (London, England)* 392.10151 (2018), pp. 951–970. ISSN: 0140-6736. DOI: [10.1016/s0140-6736\(18\)31204-2](https://doi.org/10.1016/s0140-6736(18)31204-2).
- [4] World Health Organization. *Leishmaniasis*. en. URL: <https://www.who.int/news-room/fact-sheets/detail/leishmaniasis> (visited on 10/06/2025).
- [5] World Health Organization. *Global leishmaniasis surveillance updates 2023: 3 years of the NTD road map*. en. URL: <https://www.who.int/publications/i/item/who-wer-9945-653-669> (visited on 10/08/2025).
- [6] World Health Organization. *The Global Health Observatory: Number of cases of visceral leishmaniasis reported*. World Health Organization, 2025.
- [7] François Chappuis et al. “Visceral leishmaniasis: what are the needs for diagnosis, treatment and control?” In: *Nature reviews. Microbiology* 5.11 (Nov. 2007), pp. 873–882. ISSN: 1740-1526. DOI: [10.1038/nrmicro1748](https://doi.org/10.1038/nrmicro1748).
- [8] S. P. Singh et al. “Serious underreporting of visceral leishmaniasis through passive case reporting in Bihar, India”. In: *Tropical Medicine & International Health* 11.6 (June 2006), pp. 899–905. ISSN: 1360-2276, 1365-3156. DOI: [10.1111/j.1365-3156.2006.01647.x](https://doi.org/10.1111/j.1365-3156.2006.01647.x).
- [9] Anuj Mubayi et al. “Transmission dynamics and underreporting of Kala-azar in the Indian state of Bihar”. In: *Journal of theoretical biology* 262.1 (Jan. 2010), pp. 177–185. ISSN: 0022-5193. DOI: [10.1016/j.jtbi.2009.09.012](https://doi.org/10.1016/j.jtbi.2009.09.012).
- [10] Ana Nilce Silveira Maia-Elkhoury et al. “[Analysis of visceral leishmaniasis reports by the capture-recapture method]”. In: *Revista de saude publica* 41.6 (Dec. 2007), pp. 931–937. ISSN: 0034-8910. DOI: [10.1590/s0034-89102007000600007](https://doi.org/10.1590/s0034-89102007000600007).
- [11] Jorge Alvar et al. “Leishmaniasis worldwide and global estimates of its incidence”. In: *PloS one* 7.5 (2012), e35671. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0035671](https://doi.org/10.1371/journal.pone.0035671).
- [12] World Health Organization. “Leishmaniasis in high-burden countries: an epidemiological update based on data reported in 2014”. In: *Weekly Epidemiological Record* 91.22 (June 2016). WHO Reference Number: WER No. 22, 2016, pp. 285–296.

- [13] Simon Collin et al. “Conflict and kala–azar: determinants of adverse outcomes of kala–azar among patients in southern Sudan”. In: *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 38.5 (Mar. 2004), pp. 612–619. ISSN: 1058-4838. DOI: [10.1086/381203](https://doi.org/10.1086/381203).
- [14] ME Gibson. “The identification of kala–azar and the discovery of Leishmania donovani”. In: *Medical history* 27.2 (Apr. 1983), pp. 203–213. ISSN: 0025-7273. DOI: [10.1017/s0025727300042691](https://doi.org/10.1017/s0025727300042691).
- [15] Dietmar Steverding. “The history of leishmaniasis”. In: *Parasites & vectors* 10.1 (Feb. 2017), p. 82. ISSN: 1756-3305. DOI: [10.1186/s13071-017-2028-5](https://doi.org/10.1186/s13071-017-2028-5).
- [16] Oussama Mouri et al. “Spontaneous remission of fully symptomatic visceral leishmaniasis”. In: *BMC infectious diseases* 15 (Oct. 2015), p. 445. ISSN: 1471-2334. DOI: [10.1186/s12879-015-1191-6](https://doi.org/10.1186/s12879-015-1191-6).
- [17] R Badaro et al. “New perspectives on a subclinical form of visceral leishmaniasis”. In: *The Journal of infectious diseases* 154.6 (Dec. 1986), pp. 1003–1011. ISSN: 0022-1899. DOI: [10.1093/infdis/154.6.1003](https://doi.org/10.1093/infdis/154.6.1003).
- [18] Global Burden of Disease Collaborative Network. *Global Burden of Disease Study 2023 (GBD 2023) Results*. <https://vizhub.healthdata.org/gbd-results/>. Available from: <https://vizhub.healthdata.org/gbd-results/>. Seattle, United States, 2025.
- [19] Mohammad Akhouni et al. “A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies”. In: *PLoS neglected tropical diseases* 10.3 (Mar. 2016), e0004349. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0004349](https://doi.org/10.1371/journal.pntd.0004349).
- [20] J Alvar et al. “Leishmania and human immunodeficiency virus coinfection: the first 10 years”. In: *Clinical microbiology reviews* 10.2 (Apr. 1997), pp. 298–319. ISSN: 0893-8512. DOI: [10.1128/cmr.10.2.298](https://doi.org/10.1128/cmr.10.2.298).
- [21] Pasquale Pagliano et al. “Visceral leishmaniasis in pregnancy: a case series and a systematic review of the literature.” eng. In: *The Journal of antimicrobial chemotherapy* 55.2 (Feb. 2005). Place: England, pp. 229–233. DOI: [10.1093/jac/dkh538](https://doi.org/10.1093/jac/dkh538).
- [22] Diego L Guedes et al. “Sexual Transmission of Visceral Leishmaniasis: A Neglected Story”. In: *Trends in parasitology* 36.12 (Dec. 2020), pp. 950–952. ISSN: 1471-4922. DOI: [10.1016/j.pt.2020.08.002](https://doi.org/10.1016/j.pt.2020.08.002).
- [23] WS Symmers. “Leishmaniasis acquired by contagion: a case of marital infection in Britain”. In: *Lancet (London, England)* 1.7116 (Jan. 1960), pp. 127–132. ISSN: 0140-6736. DOI: [10.1016/s0140-6736\(60\)90052-0](https://doi.org/10.1016/s0140-6736(60)90052-0).
- [24] M Magdalena Alcover et al. “Wild mammals as potential silent reservoirs of Leishmania infantum in a Mediterranean area”. In: *Preventive veterinary medicine* 175 (Feb. 2020), p. 104874. ISSN: 0167-5877. DOI: [10.1016/j.prevmed.2019.104874](https://doi.org/10.1016/j.prevmed.2019.104874).
- [25] Fabiana Raquel Ratzlaff et al. “Identification of infection by Leishmania spp. in wild and domestic animals in Brazil: a systematic review with meta-analysis (2001-2021)”. In: *Parasitology research* 122.7 (July 2023), pp. 1605–1619. ISSN: 0932-0113. DOI: [10.1007/s00436-023-07862-y](https://doi.org/10.1007/s00436-023-07862-y).

- [26] R Molina et al. "The hare (*Lepus granatensis*) as potential sylvatic reservoir of *Leishmania infantum* in Spain". In: *Veterinary parasitology* 190.1-2 (Nov. 2012), pp. 268–271. ISSN: 0304-4017. DOI: [10.1016/j.vetpar.2012.05.006](https://doi.org/10.1016/j.vetpar.2012.05.006).
- [27] Om Prakash Singh et al. "Xenodiagnosis to evaluate the infectiousness of humans to sandflies in an area endemic for visceral leishmaniasis in Bihar, India: a transmission-dynamics study". In: *The Lancet. Microbe* 2.1 (Jan. 2021), e23–e31. ISSN: 2666-5247. DOI: [10.1016/s2666-5247\(20\)30166-x](https://doi.org/10.1016/s2666-5247(20)30166-x).
- [28] Dinesh Mondal et al. "Quantifying the Infectiousness of Post-Kala-Azar Dermal Leishmaniasis Toward Sand Flies". In: *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 69.2 (July 2019), pp. 251–258. ISSN: 1058-4838. DOI: [10.1093/cid/ciy891](https://doi.org/10.1093/cid/ciy891).
- [29] Anurag Kumar Kushwaha et al. "Dogs as Reservoirs for *Leishmania donovani*, Bihar, India, 2018–2022". In: *Emerging infectious diseases* 30.12 (Dec. 2024), pp. 2604–2613. ISSN: 1080-6040. DOI: [10.3201/eid3012.240649](https://doi.org/10.3201/eid3012.240649).
- [30] Caitlin M Jones and Susan C Welburn. "Leishmaniasis Beyond East Africa". In: *Frontiers in veterinary science* 8 (2021), p. 618766. ISSN: 2297-1769. DOI: [10.3389/fvets.2021.618766](https://doi.org/10.3389/fvets.2021.618766).
- [31] Milena Menezes Corrêa Pederiva et al. "Asymptomatic *Leishmania* infection in humans: A systematic review". In: *Journal of infection and public health* 16.2 (Feb. 2023), pp. 286–294. ISSN: 1876-0341. DOI: [10.1016/j.jiph.2022.12.021](https://doi.org/10.1016/j.jiph.2022.12.021).
- [32] Gouri S Bhunia et al. "Influence of topography on the endemicity of Kala-azar: a study based on remote sensing and geographical information system". In: *Geospatial health* 4.2 (May 2010), pp. 155–165. ISSN: 1827-1987. DOI: [10.4081/gh.2010.197](https://doi.org/10.4081/gh.2010.197).
- [33] Caryn Bern et al. "Of cattle, sand flies and men: a systematic review of risk factor analyses for South Asian visceral leishmaniasis and implications for elimination". In: *PLoS neglected tropical diseases* 4.2 (Feb. 2010), e599. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0000599](https://doi.org/10.1371/journal.pntd.0000599).
- [34] Abebe Kassa Geto et al. "Prevalence of human visceral leishmaniasis and its risk factors in Eastern Africa: a systematic review and meta-analysis". In: *Frontiers in public health* 12 (2024), p. 1488741. ISSN: 2296-2565. DOI: [10.3389/fpubh.2024.1488741](https://doi.org/10.3389/fpubh.2024.1488741).
- [35] Begoña Monge-Maillo et al. "Visceral leishmaniasis and HIV coinfection in the Mediterranean region". In: *PLoS neglected tropical diseases* 8.8 (Aug. 2014), e3021. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0003021](https://doi.org/10.1371/journal.pntd.0003021).
- [36] Jorge Alvar et al. "The relationship between leishmaniasis and AIDS: the second 10 years". In: *Clinical microbiology reviews* 21.2 (Apr. 2008), 334–59, table of contents. ISSN: 0893-8512. DOI: [10.1128/cmr.00061-07](https://doi.org/10.1128/cmr.00061-07).
- [37] Organización Panamericana de la Salud. *Leishmaniasis: informe epidemiológico de las Américas. Núm. 13, diciembre 2024*. Tech. rep. Internet; 11 p. [cited 2025-10-year month day]. Disponible en: <https://iris.paho.org/handle/10665.2/63164>. Washington, D.C.: PAHO / OPS, 2024.

- [38] National Center for Vector Borne Diseases Control, Directorate General of Health Services, Ministry of Health and Family Welfare, Government of India. *Kala-Azar Situation in India*. Accessed: 13 October 2025. URL: <https://ncvbdc.mohfw.gov.in/index1.php?lang=1%5C&level=2%5C&sublinkid=5945%5C&lid=3750>.
- [39] Ermias Diro et al. “Visceral Leishmaniasis and HIV coinfection in East Africa”. In: *PLoS neglected tropical diseases* 8.6 (June 2014), e2869. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0002869](https://doi.org/10.1371/journal.pntd.0002869).
- [40] EE Zijlstra et al. “Kala-azar in displaced people from southern Sudan: epidemiological, clinical and therapeutic findings”. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 85.3 (1991), pp. 365–369. ISSN: 0035-9203. DOI: [10.1016/0035-9203\(91\)90293-8](https://doi.org/10.1016/0035-9203(91)90293-8).
- [41] Waleed Al-Salem et al. “A review of visceral leishmaniasis during the conflict in South Sudan and the consequences for East African countries”. In: *Parasites & vectors* 9 (Aug. 2016), p. 460. ISSN: 1756-3305. DOI: [10.1186/s13071-016-1743-7](https://doi.org/10.1186/s13071-016-1743-7).
- [42] Simona Stäger et al. “Immune evasive mechanisms contributing to persistent Leishmania donovani infection”. In: *Immunologic research* 47.1-3 (July 2010), pp. 14–24. ISSN: 0257-277X. DOI: [10.1007/s12026-009-8135-4](https://doi.org/10.1007/s12026-009-8135-4).
- [43] Paul M Kaye et al. “Leishmaniasis immunopathology-impact on design and use of vaccines, diagnostics and drugs”. In: *Seminars in immunopathology* 42.3 (June 2020), pp. 247–264. ISSN: 1863-2297. DOI: [10.1007/s00281-020-00788-y](https://doi.org/10.1007/s00281-020-00788-y).
- [44] Susanne Nylén and Shalini Gautam. “Immunological perspectives of leishmaniasis”. In: *Journal of global infectious diseases* 2.2 (May 2010), pp. 135–146. ISSN: 0974-777X. DOI: [10.4103/0974-777x.62876](https://doi.org/10.4103/0974-777x.62876).
- [45] GL Werneck et al. “The burden of Leishmania chagasi infection during an urban outbreak of visceral leishmaniasis in Brazil”. In: *Acta tropica* 83.1 (July 2002), pp. 13–18. ISSN: 0001-706X. DOI: [10.1016/s0001-706x\(02\)00058-x](https://doi.org/10.1016/s0001-706x(02)00058-x).
- [46] Paul Ready. “Epidemiology of visceral leishmaniasis”. en. In: *Clinical Epidemiology* (May 2014), p. 147. ISSN: 1179-1349. DOI: [10.2147/CLEP.S44267](https://doi.org/10.2147/CLEP.S44267).
- [47] Neelam Varma and Shano Naseem. “Hematologic changes in visceral leishmaniasis/kala azar”. In: *Indian journal of hematology & blood transfusion: an official journal of Indian Society of Hematology and Blood Transfusion* 26.3 (Sept. 2010), pp. 78–82. ISSN: 0971-4502. DOI: [10.1007/s12288-010-0027-1](https://doi.org/10.1007/s12288-010-0027-1).
- [48] Yasuyuki Goto et al. “Prevalence, severity, and pathogeneses of anemia in visceral leishmaniasis”. In: *Parasitology research* 116.2 (Feb. 2017), pp. 457–464. ISSN: 0932-0113. DOI: [10.1007/s00436-016-5313-x](https://doi.org/10.1007/s00436-016-5313-x).
- [49] Danielle A Zacarias et al. “Causes and consequences of higher Leishmania infantum burden in patients with kala-azar: a study of 625 patients”. In: *Tropical medicine & international health: TM & IH* 22.6 (June 2017), pp. 679–687. ISSN: 1360-2276. DOI: [10.1111/tmi.12877](https://doi.org/10.1111/tmi.12877).
- [50] Joyce M Silva et al. “Bone marrow parasite burden among patients with New World kala-azar is associated with disease severity”. In: *The American journal of tropical medicine and hygiene* 90.4 (Apr. 2014), pp. 621–626. ISSN: 0002-9637. DOI: [10.4269/ajtmh.13-0376](https://doi.org/10.4269/ajtmh.13-0376).

- [51] Carlos H N Costa et al. "From Infection to Death: An Overview of the Pathogenesis of Visceral Leishmaniasis". In: *Pathogens (Basel, Switzerland)* 12.7 (July 2023), p. 969. ISSN: 2076-0817. DOI: [10.3390/pathogens12070969](https://doi.org/10.3390/pathogens12070969).
- [52] Jeffrey D. Chulay and Anthony D. M. Bryceson. "Quantitation of Amastigotes of Leishmania Donovani in Smears of Splenic Aspirates from Patients with Visceral Leishmaniasis". In: *The American Journal of Tropical Medicine and Hygiene* 32.3 (May 1983), pp. 475–479. ISSN: 0002-9637. DOI: [10.4269/ajtmh.1983.32.475](https://doi.org/10.4269/ajtmh.1983.32.475).
- [53] Luca Galluzzi et al. "Real-time PCR applications for diagnosis of leishmaniasis". In: *Parasites & vectors* 11.1 (May 2018), p. 273. ISSN: 1756-3305. DOI: [10.1186/s13071-018-2859-8](https://doi.org/10.1186/s13071-018-2859-8).
- [54] José Bryan Rihs et al. "qPCR as a Tool for the Diagnosis of Visceral and Cutaneous Leishmaniasis: A Systematic Review and Meta-Analysis". In: *Acta parasitologica* 70.1 (Jan. 2025), p. 16. ISSN: 1230-2821. DOI: [10.1007/s11686-024-00942-8](https://doi.org/10.1007/s11686-024-00942-8).
- [55] Luka Verrest et al. "Leishmania blood parasite dynamics during and after treatment of visceral leishmaniasis in Eastern Africa: A pharmacokinetic-pharmacodynamic model". In: *PLoS neglected tropical diseases* 18.4 (Apr. 2024), e0012078. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0012078](https://doi.org/10.1371/journal.pntd.0012078).
- [56] Johan van Griensven et al. "Leishmania Antigenuria to Predict Initial Treatment Failure and Relapse in Visceral Leishmaniasis/HIV Coinfected Patients: An Exploratory Study Nested Within a Clinical Trial in Ethiopia". In: *Frontiers in cellular and infection microbiology* 8 (2018), p. 94. ISSN: 2235-2988. DOI: [10.3389/fcimb.2018.00094](https://doi.org/10.3389/fcimb.2018.00094).
- [57] C Riera et al. "Evaluation of a latex agglutination test (KAtex) for detection of Leishmania antigen in urine of patients with HIV-Leishmania coinfection: value in diagnosis and post-treatment follow-up". In: *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 23.12 (Dec. 2004), pp. 899–904. ISSN: 0934-9723. DOI: [10.1007/s10096-004-1249-7](https://doi.org/10.1007/s10096-004-1249-7).
- [58] Marleen Boelaert et al. "Rapid tests for the diagnosis of visceral leishmaniasis in patients with suspected disease". In: *The Cochrane database of systematic reviews* 6 (June 2014), p. CD009135. ISSN: 1469-493X. DOI: [10.1002/14651858.cd009135.pub2](https://doi.org/10.1002/14651858.cd009135.pub2).
- [59] World Health Organization. Expert Committee on the Control of Leishmaniases. *Control of the Leishmaniases: report of a meeting of the WHO Expert Committee on the Control of Leishmaniases, Geneva, 22–26 March 2010*. Tech. rep. 949. Also published via WHO IRIS platform. Geneva: World Health Organization, 2010.
- [60] Ministry of Health, Brasil, ed. *Guia de Vigilância em Saúde, Volume 2 (6ª edição revisada)*. 6th ed. Updated 23/04/2024. Brasília, Brasil: Ministry of Health, 2024.
- [61] S Sundar. "Drug resistance in Indian visceral leishmaniasis". In: *Tropical medicine & international health: TM&IH* 6.11 (Nov. 2001), pp. 849–854. ISSN: 1360-2276. DOI: [10.1046/j.1365-3156.2001.00778.x](https://doi.org/10.1046/j.1365-3156.2001.00778.x).

- [62] Fabiana Alves et al. “Recent Development of Visceral Leishmaniasis Treatments: Successes, Pitfalls, and Perspectives”. In: *Clinical microbiology reviews* 31.4 (Oct. 2018), e00048–18. ISSN: 0893-8512. DOI: [10.1128/cmr.00048-18](https://doi.org/10.1128/cmr.00048-18).
- [63] Shyam Sundar et al. “Oral miltefosine for Indian visceral leishmaniasis”. In: *The New England journal of medicine* 347.22 (Nov. 2002), pp. 1739–1746. ISSN: 0028-4793. DOI: [10.1056/nejmoa021556](https://doi.org/10.1056/nejmoa021556).
- [64] Sujit Kumar Bhattacharya et al. “Phase 4 Trial of Miltefosine for the Treatment of Indian Visceral Leishmaniasis”. en. In: *The Journal of Infectious Diseases* 196.4 (Aug. 2007), pp. 591–598. ISSN: 0022-1899, 1537-6613. DOI: [10.1086/519690](https://doi.org/10.1086/519690).
- [65] World Health Organization. *Regional Strategic Framework for Elimination of Kala-azar from the South-East Asia Region (2005 - 2015)*. Delhi: WHO Regional Office for South-East Asia, 2005.
- [66] Suman Rijal et al. “Increasing Failure of Miltefosine in the Treatment of Kala-azar in Nepal and the Potential Role of Parasite Drug Resistance, Reinfestation, or Noncompliance”. en. In: *Clinical Infectious Diseases* 56.11 (June 2013), pp. 1530–1538. ISSN: 1537-6591, 1058-4838. DOI: [10.1093/cid/cit102](https://doi.org/10.1093/cid/cit102).
- [67] Shyam Sundar et al. “Efficacy of miltefosine in the treatment of visceral leishmaniasis in India after a decade of use”. In: *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 55.4 (Aug. 2012), pp. 543–550. ISSN: 1058-4838. DOI: [10.1093/cid/cis474](https://doi.org/10.1093/cid/cis474).
- [68] Mitali Chatterjee et al. “Challenges in sustaining the elimination of visceral leishmaniasis in India”. In: *International health* (June 2025). ISSN: 1876-3413. DOI: [10.1093/inthealth/ihaf063](https://doi.org/10.1093/inthealth/ihaf063).
- [69] Juliana B T Carnielli et al. “Natural Resistance of *Leishmania infantum* to Miltefosine Contributes to the Low Efficacy in the Treatment of Visceral Leishmaniasis in Brazil”. In: *The American journal of tropical medicine and hygiene* 101.4 (Oct. 2019), pp. 789–794. ISSN: 0002-9637. DOI: [10.4269/ajtmh.18-0949](https://doi.org/10.4269/ajtmh.18-0949).
- [70] World Health Organization. *WHO and Gilead Sciences extend collaborative agreement to enhance access to treatment for visceral leishmaniasis*. Accessed: October 17th 2025. Jan. 2023. URL: <https://www.who.int/news-room/detail/26-01-2023-who-and-gilead-sciences-extend-collaborative-agreement-to-enhance-access-to-treatment-for-visceral-leishmaniasis>.
- [71] Shyam Sundar et al. “Injectable Paromomycin for Visceral Leishmaniasis in India”. en. In: *New England Journal of Medicine* 356.25 (June 2007), pp. 2571–2581. ISSN: 0028-4793, 1533-4406. DOI: [10.1056/NEJMoa066536](https://doi.org/10.1056/NEJMoa066536).
- [72] Ahmed Musa et al. “Sodium stibogluconate (SSG) & paromomycin combination compared to SSG for visceral leishmaniasis in East Africa: a randomised controlled trial”. In: *PLoS neglected tropical diseases* 6.6 (2012), e1674. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0001674](https://doi.org/10.1371/journal.pntd.0001674).
- [73] Ahmed M Musa et al. “Paromomycin and Miltefosine Combination as an Alternative to Treat Patients With Visceral Leishmaniasis in Eastern Africa: A Randomized, Controlled, Multicountry Trial”. In: *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 76.3 (Feb. 2023), e1177–e1185. ISSN: 1058-4838. DOI: [10.1093/cid/ciac643](https://doi.org/10.1093/cid/ciac643).

- [74] Shyam Sundar et al. “Comparison of short-course multidrug treatment with standard therapy for visceral leishmaniasis in India: an open-label, non-inferiority, randomised controlled trial”. In: *Lancet (London, England)* 377.9764 (Feb. 2011), pp. 477–486. ISSN: 0140-6736. DOI: [10.1016/s0140-6736\(10\)62050-8](https://doi.org/10.1016/s0140-6736(10)62050-8).
- [75] World Health Organization. *Kala-Azar Elimination Programme: Report of a WHO Consultation of Partners, Geneva, Switzerland, 10–11 February 2015*. Tech. rep. WHO/HTM/NTD/IDM/2015.8. Geneva: World Health Organization, 2015.
- [76] World Health Organization, Regional Office for South-East Asia. *Report of Meeting of the Regional Technical Advisory Group (RTAG) on visceral leishmaniasis and the national visceral leishmaniasis programme managers of endemic member states: Virtual Meeting, 5–8 October 2020*. Tech. rep. SEA-CD-329. Meeting report; 51 pp. New Delhi: World Health Organization. Regional Office for South-East Asia, Apr. 2021.
- [77] World Health Organization. *Ending the Neglect to Attain the Sustainable Development Goals: A Road Map for Neglected Tropical Diseases 2021–2030*. Geneva: World Health Organization, 2021. ISBN: 978-92-4-001035-2.
- [78] World Health Organization. *Global report on neglected tropical diseases 2025*. Geneva: World Health Organization, 2025. ISBN: 978-92-4-011404-3.
- [79] Vidya Nand Ravi Das et al. “Repeated training of accredited social health activists (ASHAs) for improved detection of visceral leishmaniasis cases in Bihar, India”. en. In: *Pathogens and Global Health* 110.1 (Jan. 2016), pp. 33–35. ISSN: 2047-7724, 2047-7732. DOI: [10.1080/20477724.2016.1156902](https://doi.org/10.1080/20477724.2016.1156902).
- [80] Shyam Sundar et al. “Visceral leishmaniasis elimination targets in India, strategies for preventing resurgence”. In: *Expert review of anti-infective therapy* 16.11 (Nov. 2018), pp. 805–812. ISSN: 1478-7210. DOI: [10.1080/14787210.2018.1532790](https://doi.org/10.1080/14787210.2018.1532790).
- [81] Nazia Nagi. “Bangladesh eliminates visceral leishmaniasis”. In: *The Lancet. Microbe* 5.5 (May 2024), e420. ISSN: 2666-5247. DOI: [10.1016/s2666-5247\(24\)00028-4](https://doi.org/10.1016/s2666-5247(24)00028-4).
- [82] Dhruv K Pandey et al. “Kala-azar elimination in India: reflections on success and sustainability”. In: *International health* 17.4 (July 2025), pp. 416–422. ISSN: 1876-3413. DOI: [10.1093/inthealth/ihaf013](https://doi.org/10.1093/inthealth/ihaf013).
- [83] Epke A Le Rutte et al. “Elimination of visceral leishmaniasis in the Indian subcontinent: a comparison of predictions from three transmission models”. In: *Epidemics* 18 (Mar. 2017), pp. 67–80. ISSN: 1755-4365. DOI: [10.1016/j.epidem.2017.01.002](https://doi.org/10.1016/j.epidem.2017.01.002).
- [84] Jorge Alvar et al. “Towards the elimination of visceral leishmaniasis as a public health problem in east Africa: reflections on an enhanced control strategy and a call for action”. In: *The Lancet. Global health* 9.12 (Dec. 2021), e1763–e1769. ISSN: 2214-109X. DOI: [10.1016/s2214-109x\(21\)00392-2](https://doi.org/10.1016/s2214-109x(21)00392-2).
- [85] World Health Organization. *Strategic framework for the elimination of visceral leishmaniasis as a public health problem in eastern Africa 2023–2030: web annex: the Nairobi Declaration*. Tech. rep. WHO Reference Number: B09041. World Health Organization, June 2024, p. 5.

- [86] World Health Organization. *Strategic framework for the elimination of visceral leishmaniasis as a public health problem in eastern Africa 2023–2030*. Geneva: World Health Organization, 2024, p. 82. ISBN: 978-92-4-009420-8.
- [87] World Health Organization, Regional Office for South-East Asia and WHO Team (Control of Neglected Tropical Diseases) TDR (Special Programme for Research and Training in Tropical Diseases). *Indicators for monitoring and evaluation of the kala-azar elimination programme*. Technical Document ISBN 978 92 4 150037 1. Bangladesh, India and Nepal - August 2010. 35 pp. New Delhi, India: World Health Organization, Regional Office for South-East Asia, 2010.
- [88] World Health Organization. *Target product profile for a diagnostic test to confirm cure of visceral leishmaniasis*. Tech. rep. ISBN 978-92-4-009181-8. World Health Organization, June 2024, p. 16.
- [89] Rutuja Chhajed et al. “Estimating the proportion of relapse following treatment of Visceral Leishmaniasis: meta-analysis using Infectious Diseases Data Observatory (IDDO) systematic review”. en. In: *The Lancet Regional Health - Southeast Asia* 22 (Mar. 2024), p. 100317. ISSN: 27723682. DOI: [10.1016/j.lansea.2023.100317](https://doi.org/10.1016/j.lansea.2023.100317).
- [90] Prabin Dahal et al. “Design, Conduct, Analysis, and Reporting of Therapeutic Efficacy Studies in Visceral Leishmaniasis: A Systematic Review of Published Reports, 2000-2021”. In: *The American journal of tropical medicine and hygiene* 111.2 (Aug. 2024), pp. 365–376. ISSN: 0002-9637. DOI: [10.4269/ajtmh.23-0458](https://doi.org/10.4269/ajtmh.23-0458).
- [91] Robert Kimutai et al. “Safety and Effectiveness of Sodium Stibogluconate and Paromomycin Combination for the Treatment of Visceral Leishmaniasis in Eastern Africa: Results from a Pharmacovigilance Programme”. In: *Clinical drug investigation* 37.3 (Mar. 2017), pp. 259–272. ISSN: 1173-2563. DOI: [10.1007/s40261-016-0481-0](https://doi.org/10.1007/s40261-016-0481-0).
- [92] Atia M Atia et al. “Sodium stibogluconate and paromomycin for treating visceral leishmaniasis under routine conditions in eastern Sudan”. In: *Tropical medicine & international health: TM&IH* 20.12 (Dec. 2015), pp. 1674–1684. ISSN: 1360-2276. DOI: [10.1111/tmi.12603](https://doi.org/10.1111/tmi.12603).
- [93] Yosef Melaku et al. “Treatment of kala-azar in southern Sudan using a 17-day regimen of sodium stibogluconate combined with paromomycin: a retrospective comparison with 30-day sodium stibogluconate monotherapy”. In: *The American journal of tropical medicine and hygiene* 77.1 (July 2007), pp. 89–94. ISSN: 0002-9637. DOI: [10.4269/ajtmh.2007.77.89](https://doi.org/10.4269/ajtmh.2007.77.89).
- [94] Shyam Sundar et al. “Effectiveness of Single-Dose Liposomal Amphotericin B in Visceral Leishmaniasis in Bihar”. en. In: *The American Journal of Tropical Medicine and Hygiene* 101.4 (Oct. 2019), pp. 795–798. ISSN: 0002-9637, 1476-1645. DOI: [10.4269/ajtmh.19-0179](https://doi.org/10.4269/ajtmh.19-0179).
- [95] Sakib Burza et al. “One-year follow-up of immunocompetent male patients treated with miltefosine for primary visceral leishmaniasis in Bihar, India”. In: *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 57.9 (Nov. 2013), pp. 1363–1364. ISSN: 1058-4838. DOI: [10.1093/cid/cit508](https://doi.org/10.1093/cid/cit508).

- [96] Sakib Burza et al. "Risk factors for visceral leishmaniasis relapse in immunocompetent patients following treatment with 20 mg/kg liposomal amphotericin B (Ampicin) in Bihar, India". eng. In: *PLoS neglected tropical diseases* 8.1 (2014), e2536. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0002536](https://doi.org/10.1371/journal.pntd.0002536).
- [97] Vishal Goyal et al. "Field effectiveness of new visceral leishmaniasis regimens after 1 year following treatment within public health facilities in Bihar, India". In: *PLoS neglected tropical diseases* 13.9 (Sept. 2019), e0007726. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0007726](https://doi.org/10.1371/journal.pntd.0007726).
- [98] World Health Organization. Global Malaria Programme. *Methods for surveillance of antimalarial drug efficacy*. ISBN: 978 92 4 159753 1. World Health Organization. Geneva, Oct. 2009.
- [99] Susanne U Franssen et al. "Diversity and Within-Host Evolution of Leishmania donovani from Visceral Leishmaniasis Patients with and without HIV Coinfection in Northern Ethiopia". In: *mBio* 12.3 (June 2021), e0097121. ISSN: 2150-7511. DOI: [10.1128/mbio.00971-21](https://doi.org/10.1128/mbio.00971-21).
- [100] Miguel A Morales et al. "Relapses versus reinfections in patients coinfected with Leishmania infantum and human immunodeficiency virus type 1". In: *The Journal of infectious diseases* 185.10 (May 2002), pp. 1533–1537. ISSN: 0022-1899. DOI: [10.1086/340219](https://doi.org/10.1086/340219).
- [101] Tesfaye Gelanew et al. "Inference of population structure of Leishmania donovani strains isolated from different Ethiopian visceral leishmaniasis endemic areas". In: *PLoS neglected tropical diseases* 4.11 (Nov. 2010), e889. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0000889](https://doi.org/10.1371/journal.pntd.0000889).
- [102] Laurence Lachaud et al. "Parasite susceptibility to amphotericin B in failures of treatment for visceral leishmaniasis in patients coinfected with HIV type 1 and Leishmania infantum". In: *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 48.2 (Jan. 2009), e16–22. ISSN: 1058-4838. DOI: [10.1086/595710](https://doi.org/10.1086/595710).
- [103] Gary S Collins et al. "TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods". en. In: *BMJ* (Apr. 2024), e078378. ISSN: 1756-1833. DOI: [10.1136/bmj-2023-078378](https://doi.org/10.1136/bmj-2023-078378).
- [104] Siddhivinayak Hirve et al. "Transmission Dynamics of Visceral Leishmaniasis in the Indian subcontinent - A Systematic Literature Review". en. In: *PLOS Neglected Tropical Diseases* 10.8 (Aug. 2016). Ed. by Mitali Chatterjee, e0004896. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0004896](https://doi.org/10.1371/journal.pntd.0004896).
- [105] Gustavo De Almeida Santos et al. "Systematic Review of Treatment Failure and Clinical Relapses in Leishmaniasis from a Multifactorial Perspective: Clinical Aspects, Factors Associated with the Parasite and Host". en. In: *Tropical Medicine and Infectious Disease* 8.9 (Aug. 2023), p. 430. DOI: [10.3390/tropicalmed8090430](https://doi.org/10.3390/tropicalmed8090430).
- [106] Gláucia F Cota et al. "Predictors of visceral leishmaniasis relapse in HIV-infected patients: a systematic review". In: *PLoS neglected tropical diseases* 5.6 (June 2011), e1153. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0001153](https://doi.org/10.1371/journal.pntd.0001153).

- [107] Mekuriaw Alemayehu et al. "Magnitude of visceral leishmaniasis and poor treatment outcome among HIV patients: meta-analysis and systematic review". In: *HIV/AIDS (Auckland, N.Z.)* 8 (2016), pp. 75–81. ISSN: 1179-1373. DOI: [10.2147/hiv.s96883](https://doi.org/10.2147/hiv.s96883).
- [108] Vishal Goyal et al. "Long-term incidence of relapse and post-kala-azar dermal leishmaniasis after three different visceral leishmaniasis treatment regimens in Bihar, India". en. In: *PLOS Neglected Tropical Diseases* 14.7 (July 2020). Ed. by Guilherme L. Werneck, e0008429. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0008429](https://doi.org/10.1371/journal.pntd.0008429).
- [109] Bart Ostyn et al. "Failure of Miltefosine Treatment for Visceral Leishmaniasis in Children and Men in South-East Asia". en. In: *PLoS ONE* 9.6 (June 2014). Ed. by Lisa Fp. Ng, e100220. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0100220](https://doi.org/10.1371/journal.pone.0100220).
- [110] Emiliano Lucero et al. "Effectiveness and Safety of Short Course Liposomal Amphotericin B (AmBisome) as First Line Treatment for Visceral Leishmaniasis in Bangladesh". en. In: *PLOS Neglected Tropical Diseases* 9.4 (Apr. 2015). Ed. by Marleen Boelaert, e0003699. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0003699](https://doi.org/10.1371/journal.pntd.0003699).
- [111] Dinesh Mondal et al. "Relationship between treatment regimens for visceral leishmaniasis and development of post-kala-azar dermal leishmaniasis and visceral leishmaniasis relapse: A cohort study from Bangladesh". en. In: *PLOS Neglected Tropical Diseases* 13.8 (Aug. 2019). Ed. by Jesus G. Valenzuela, e0007653. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0007653](https://doi.org/10.1371/journal.pntd.0007653).
- [112] Stanislaw Gorski et al. "Visceral leishmaniasis relapse in Southern Sudan (1999-2007): a retrospective study of risk factors and trends". In: *PLoS neglected tropical diseases* 4.6 (June 2010), e705. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0000705](https://doi.org/10.1371/journal.pntd.0000705).
- [113] Grace C Kennedy et al. "Visceral leishmaniasis follow-up and treatment outcomes in Tiaty East and West sub-counties, Kenya: Cure, relapse, and Post Kala-azar Dermal Leishmaniasis". In: *PLoS One* 19.6 (2024), e0306067. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0306067](https://doi.org/10.1371/journal.pone.0306067).
- [114] Maia Kajaia et al. "Risk factors for relapse of visceral leishmaniasis in Georgia". In: *Tropical medicine & international health: TM & IH* 16.2 (Feb. 2011), pp. 186–192. ISSN: 1360-2276. DOI: [10.1111/j.1365-3156.2010.02694.x](https://doi.org/10.1111/j.1365-3156.2010.02694.x).
- [115] KhO Meliia and OP Zenaishvili. "[Ultrasound study of parenchymatous organs in visceral leishmaniasis]". In: *Meditinskaiia parazitologiiia i parazitarnye bolezni* 2 (2006), pp. 31–33. ISSN: 0025-8326.
- [116] Ermias Diro et al. "Long term outcomes and prognostics of visceral leishmaniasis in HIV infected patients with use of pentamidine as secondary prophylaxis based on CD4 level: a prospective cohort study in Ethiopia". In: *PLoS neglected tropical diseases* 13.2 (Feb. 2019), e0007132. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0007132](https://doi.org/10.1371/journal.pntd.0007132).
- [117] Sakib Burza et al. "Visceral leishmaniasis and HIV co-infection in Bihar, India: long-term effectiveness and treatment outcomes with liposomal amphotericin B (AmBisome)". In: *PLoS neglected tropical diseases* 8.8 (Aug. 2014), e3053. ISSN: 1935-2727. DOI: [10.1371/journal.pntd.0003053](https://doi.org/10.1371/journal.pntd.0003053).

- [118] Charles Abongomera et al. “The Risk and Predictors of Visceral Leishmaniasis Relapse in Human Immunodeficiency Virus-Coinfected Patients in Ethiopia: A Retrospective Cohort Study”. In: *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 65.10 (Oct. 2017), pp. 1703–1710. ISSN: 1058-4838. DOI: [10.1093/cid/cix607](https://doi.org/10.1093/cid/cix607).
- [119] Larissa D L N Costa et al. “Factors associated with relapse and hospital death in patients coinfected with visceral leishmaniasis and HIV: a longitudinal study”. In: *BMC infectious diseases* 23.1 (Mar. 2023), p. 141. ISSN: 1471-2334. DOI: [10.1186/s12879-023-08009-1](https://doi.org/10.1186/s12879-023-08009-1).
- [120] Johan Van Griensven et al. “The status of combination therapy for visceral leishmaniasis: an updated review”. en. In: *The Lancet Infectious Diseases* 24.1 (2024), e36–e46. DOI: [10.1016/S1473-3099\(23\)00353-5](https://doi.org/10.1016/S1473-3099(23)00353-5).
- [121] Eltahir A. G. Khalil et al. “Safety and Efficacy of Single Dose versus Multiple Doses of AmBisome for Treatment of Visceral Leishmaniasis in Eastern Africa: A Randomised Trial”. en. In: *PLoS Neglected Tropical Diseases* 8.1 (Jan. 2014). Ed. by Elodie Ghedin, e2613. DOI: [10.1371/journal.pntd.0002613](https://doi.org/10.1371/journal.pntd.0002613).
- [122] Ahmed M. Musa et al. “Paromomycin for the Treatment of Visceral Leishmaniasis in Sudan: A Randomized, Open-Label, Dose-Finding Study”. en. In: *PLoS Neglected Tropical Diseases* 4.10 (Oct. 2010). Ed. by Diana N. J. Lockwood, e855. DOI: [10.1371/journal.pntd.0000855](https://doi.org/10.1371/journal.pntd.0000855).
- [123] Asrat Hailu et al. “Geographical Variation in the Response of Visceral Leishmaniasis to Paromomycin in East Africa: A Multicentre, Open-Label, Randomized Trial”. en. In: *PLoS Neglected Tropical Diseases* 4.10 (Oct. 2010). Ed. by Diana N. J. Lockwood, e709. DOI: [10.1371/journal.pntd.0000709](https://doi.org/10.1371/journal.pntd.0000709).
- [124] Thomas P. C. Dorlo et al. “Optimal Dosing of Miltefosine in Children and Adults with Visceral Leishmaniasis”. en. In: *Antimicrobial Agents and Chemotherapy* 56.7 (2012), pp. 3864–3872. DOI: [10.1128/AAC.00292-12](https://doi.org/10.1128/AAC.00292-12).
- [125] Thomas P C Dorlo et al. “Visceral leishmaniasis relapse hazard is linked to reduced miltefosine exposure in patients from Eastern Africa: a population pharmacokinetic/pharmacodynamic study”. en. In: *Journal of Antimicrobial Chemotherapy* 72.11 (Nov. 2017), pp. 3131–3140. ISSN: 0305-7453, 1460-2091. DOI: [10.1093/jac/dkx283](https://doi.org/10.1093/jac/dkx283).
- [126] Jane Mbui et al. “Pharmacokinetics, Safety, and Efficacy of an Allometric Miltefosine Regimen for the Treatment of Visceral Leishmaniasis in Eastern African Children: An Open-label, Phase II Clinical Trial”. en. In: *Clinical Infectious Diseases* 68.9 (Apr. 2019), pp. 1530–1538. DOI: [10.1093/cid/ciy747](https://doi.org/10.1093/cid/ciy747).
- [127] Piero L Olliaro et al. “Treatment options for visceral leishmaniasis: a systematic review of clinical studies done in India, 1980-2004”. In: *The Lancet. Infectious diseases* 5.12 (Dec. 2005), pp. 763–774. ISSN: 1473-3099. DOI: [10.1016/s1473-3099\(05\)70296-6](https://doi.org/10.1016/s1473-3099(05)70296-6).
- [128] Fakhri Jeddi et al. “Antimony Resistance in Leishmania , Focusing on Experimental Research”. en. In: *Journal of Tropical Medicine* 2011 (2011), pp. 1–15. DOI: [10.1155/2011/695382](https://doi.org/10.1155/2011/695382).

- [129] Manu Vanaerschot et al. “Antimonial resistance in Leishmania donovani is associated with increased in vivo parasite burden”. In: *PLoS one* 6.8 (2011), e23120. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0023120](https://doi.org/10.1371/journal.pone.0023120).
- [130] R. Singh et al. “Decreased miltefosine susceptibility in clinical isolates of Leishmania donovani derived from visceral leishmaniasis and post kala–azar dermal leishmaniasis: Apparent mechanisms and clinical implications”. en. In: *International Journal of Infectious Diseases* 45 (2016), p. 381. DOI: [10.1016/j.ijid.2016.02.817](https://doi.org/10.1016/j.ijid.2016.02.817).
- [131] Keshav Rai et al. “Relapse after treatment with miltefosine for visceral leishmaniasis is associated with increased infectivity of the infecting Leishmania donovani strain”. In: *mBio* 4.5 (Oct. 2013), e00611–13. ISSN: 2150-7511. DOI: [10.1128/mbio.00611-13](https://doi.org/10.1128/mbio.00611-13).
- [132] Gabriel Naylor-Leyland et al. “The increasing incidence of visceral leishmaniasis relapse in South Sudan: A retrospective analysis of field patient data from 2001–2018”. en. In: *PLOS Neglected Tropical Diseases* 16.8 (Aug. 2022). Ed. by Mitali Chatterjee, e0010696. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0010696](https://doi.org/10.1371/journal.pntd.0010696).
- [133] Anke E. Kip et al. “Systematic Review of Biomarkers To Monitor Therapeutic Response in Leishmaniasis”. en. In: *Antimicrobial Agents and Chemotherapy* 59.1 (2015), pp. 1–14. DOI: [10.1128/AAC.04298-14](https://doi.org/10.1128/AAC.04298-14).
- [134] Spinello Antinori et al. “Clinical use of polymerase chain reaction performed on peripheral blood and bone marrow samples for the diagnosis and monitoring of visceral leishmaniasis in HIV-infected and HIV-uninfected patients: a single-center, 8-year experience in Italy and review of the literature”. eng. In: *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 44.12 (2007), pp. 1602–1610. DOI: [10.1086/518167](https://doi.org/10.1086/518167).
- [135] Nathalie Bourgeois et al. “Long-term monitoring of visceral leishmaniasis in patients with AIDS: relapse risk factors, value of polymerase chain reaction, and potential impact on secondary prophylaxis”. eng. In: *Journal of Acquired Immune Deficiency Syndromes (1999)* 48.1 (May 2008), pp. 13–19. DOI: [10.1097/QAI.0b013e318166af5d](https://doi.org/10.1097/QAI.0b013e318166af5d).
- [136] Gláucia Fernandes Cota et al. “Exploring prognosis in chronic relapsing visceral leishmaniasis among HIV-infected patients: Circulating Leishmania DNA”. eng. In: *Acta Tropica* 172 (Aug. 2017), pp. 186–191. DOI: [10.1016/j.actatropica.2017.05.011](https://doi.org/10.1016/j.actatropica.2017.05.011).
- [137] Israel Molina et al. “Ultrasensitive real-time PCR for the clinical management of visceral leishmaniasis in HIV-Infected patients”. eng. In: *The American Journal of Tropical Medicine and Hygiene* 89.1 (2013), pp. 105–110. DOI: [10.4269/ajtmh.12-0527](https://doi.org/10.4269/ajtmh.12-0527).
- [138] Antonio Carlos Nicodemo et al. “Usefulness of kDNA PCR in the diagnosis of visceral leishmaniasis reactivation in co-infected patients”. eng. In: *Revista Do Instituto De Medicina Tropical De Sao Paulo* 55.6 (2013), pp. 4290–431. DOI: [10.1590/S0036-46652013000600011](https://doi.org/10.1590/S0036-46652013000600011).

- [139] Sandeep Verma et al. “Development of a rapid loop-mediated isothermal amplification assay for diagnosis and assessment of cure of Leishmania infection”. eng. In: *BMC infectious diseases* 17.1 (Mar. 2017), p. 223. DOI: [10.1186/s12879-017-2318-8](https://doi.org/10.1186/s12879-017-2318-8).
- [140] Faria Hossain et al. “Real-time PCR in detection and quantitation of Leishmania donovani for the diagnosis of Visceral Leishmaniasis patients and the monitoring of their response to treatment”. In: *PLoS one* 12.9 (2017), e0185606. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0185606](https://doi.org/10.1371/journal.pone.0185606).
- [141] Charles Mary et al. “Reference values for Leishmania infantum parasitemia in different clinical presentations: quantitative polymerase chain reaction for therapeutic monitoring and patient follow-up”. eng. In: *The American Journal of Tropical Medicine & Hygiene* 75.5 (Nov. 2006), pp. 858–863.
- [142] Medhavi Sudarshan et al. “Study of parasite kinetics with antileishmanial drugs using real-time quantitative PCR in Indian visceral leishmaniasis”. en. In: *Journal of Antimicrobial Chemotherapy* 66.8 (2011), pp. 1751–1755. DOI: [10.1093/jac/dkr185](https://doi.org/10.1093/jac/dkr185).
- [143] Luka Verrest et al. “Blood Parasite Load as an Early Marker to Predict Treatment Response in Visceral Leishmaniasis in Eastern Africa”. In: *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 73.5 (Sept. 2021), pp. 775–782. ISSN: 1058-4838. DOI: [10.1093/cid/ciab124](https://doi.org/10.1093/cid/ciab124).
- [144] Tapan Bhattacharyya et al. “IgG1 as a potential biomarker of post-chemotherapeutic relapse in visceral leishmaniasis, and adaptation to a rapid diagnostic test”. eng. In: *PLoS neglected tropical diseases* 8.10 (Oct. 2014), e3273. DOI: [10.1371/journal.pntd.0003273](https://doi.org/10.1371/journal.pntd.0003273).
- [145] Dinesh Mondal et al. “Relationship of Serum Antileishmanial Antibody With Development of Visceral Leishmaniasis, Post-kala-azar Dermal Leishmaniasis and Visceral Leishmaniasis Relapse”. In: *Frontiers in Microbiology* 10 (Oct. 2019), p. 2268. DOI: [10.3389/fmicb.2019.02268](https://doi.org/10.3389/fmicb.2019.02268).
- [146] Guy Mollett et al. “Detection of Immunoglobulin G1 Against rK39 Improves Monitoring of Treatment Outcomes in Visceral Leishmaniasis”. en. In: *Clinical Infectious Diseases* 69.7 (2019), pp. 1130–1135. DOI: [10.1093/cid/ciy1062](https://doi.org/10.1093/cid/ciy1062).
- [147] Renata Caetano Kuschnir et al. “High levels of anti-Leishmania IgG3 and low CD4+ T cells count were associated with relapses in visceral leishmaniasis”. en. In: *BMC Infectious Diseases* 21.1 (2021), p. 369. DOI: [10.1186/s12879-021-06051-5](https://doi.org/10.1186/s12879-021-06051-5).
- [148] Lloyd A. C. Chapman et al. “Quantification of the natural history of visceral leishmaniasis and consequences for control”. eng. In: *Parasites & Vectors* 8 (Oct. 2015), p. 521. DOI: [10.1186/s13071-015-1136-3](https://doi.org/10.1186/s13071-015-1136-3).
- [149] Johan van Griensven et al. “Prediction of visceral leishmaniasis development in a highly exposed HIV cohort in Ethiopia based on Leishmania infection markers: results from the PreLeisH study”. eng. In: *EBioMedicine* 110 (Dec. 2024), p. 105474. DOI: [10.1016/j.ebiom.2024.105474](https://doi.org/10.1016/j.ebiom.2024.105474).
- [150] Mukesh Samant et al. “Role of Cytokines in Experimental and Human Visceral Leishmaniasis”. eng. In: *Frontiers in Cellular and Infection Microbiology* 11 (2021), p. 624009. DOI: [10.3389/fcimb.2021.624009](https://doi.org/10.3389/fcimb.2021.624009).

- [151] Yegnasew Takele et al. “Immunological factors, but not clinical features, predict visceral leishmaniasis relapse in patients co-infected with HIV”. In: *Cell reports. Medicine* 3.1 (Jan. 2022), p. 100487. ISSN: 2666-3791. DOI: [10.1016/j.xcrm.2021.100487](https://doi.org/10.1016/j.xcrm.2021.100487).
- [152] Dagimawie Tadesse et al. “Antibody and cytokine levels in visceral leishmaniasis patients with varied parasitemia before, during, and after treatment in patients admitted to Arba Minch General Hospital, southern Ethiopia”. en. In: *PLOS Neglected Tropical Diseases* 15.8 (Aug. 2021). Ed. by Susan M. Bueno, e0009632. DOI: [10.1371/journal.pntd.0009632](https://doi.org/10.1371/journal.pntd.0009632).
- [153] Diego Lins Guedes et al. “Comparison of serum cytokine levels in symptomatic and asymptomatic HIV-Leishmania coinfected individuals from a Brazilian visceral leishmaniasis endemic area”. en. In: *PLOS Neglected Tropical Diseases* 16.6 (2022). Ed. by Ikram Guizani, e0010542. DOI: [10.1371/journal.pntd.0010542](https://doi.org/10.1371/journal.pntd.0010542).
- [154] Anke E Kip et al. “Macrophage Activation Marker Neopterin: A Candidate Biomarker for Treatment Response and Relapse in Visceral Leishmaniasis”. In: *Frontiers in cellular and infection microbiology* 8 (2018), p. 181. ISSN: 2235-2988. DOI: [10.3389/fcimb.2018.00181](https://doi.org/10.3389/fcimb.2018.00181).
- [155] Ana Torres et al. “Proteomics of plasma-derived extracellular vesicles from human patients identifies biomarkers for monitoring visceral leishmaniasis therapy”. In: *Frontiers in Immunology* 16 (2025), p. 1646335. DOI: [10.3389/fimmu.2025.1646335](https://doi.org/10.3389/fimmu.2025.1646335).
- [156] James Wilson et al. “Prognostic Prediction Models for Clinical Outcomes in Patients Diagnosed with Visceral Leishmaniasis: Protocol for a Systematic Review”. In: *BMJ open* 13.10 (Oct. 24, 2023), e075597. ISSN: 2044-6055. DOI: [10.1136/bmjopen-2023-075597](https://doi.org/10.1136/bmjopen-2023-075597). PMID: 37879686.
- [157] K. G M Moons et al. “Prognosis and prognostic research: what, why, and how?” en. In: *BMJ* 338.feb23 1 (Feb. 2009), b375–b375. ISSN: 0959-8138, 1468-5833. DOI: [10.1136/bmj.b375](https://doi.org/10.1136/bmj.b375).
- [158] Karel G.M. Moons et al. “Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration”. en. In: *Annals of Internal Medicine* 162.1 (Jan. 2015), W1–W73. ISSN: 0003-4819, 1539-3704. DOI: [10.7326/M14-0698](https://doi.org/10.7326/M14-0698).
- [159] Laure Wynants et al. “Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal”. en. In: *BMJ* (Apr. 2020), p. m1328. ISSN: 1756-1833. DOI: [10.1136/bmj.m1328](https://doi.org/10.1136/bmj.m1328).
- [160] Lauren S. Peetluk et al. “Systematic review of prediction models for pulmonary tuberculosis treatment outcomes in adults”. en. In: *BMJ Open* 11.3 (Mar. 2021), e044687. ISSN: 2044-6055, 2044-6055. DOI: [10.1136/bmjopen-2020-044687](https://doi.org/10.1136/bmjopen-2020-044687).
- [161] Tsi Njim and Bayee Swiri Tanyitiku. “Prognostic models for the clinical management of malaria and its complications: a systematic review”. en. In: *BMJ Open* 9.11 (Nov. 2019), e030793. ISSN: 2044-6055, 2044-6055. DOI: [10.1136/bmjopen-2019-030793](https://doi.org/10.1136/bmjopen-2019-030793).

- [162] Ewout W. Steyerberg et al. "Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research". en. In: *PLoS Medicine* 10.2 (Feb. 2013), e1001381. ISSN: 1549-1676. DOI: [10.1371/journal.pmed.1001381](https://doi.org/10.1371/journal.pmed.1001381).
- [163] Charles Abongomera et al. "Development and External Validation of a Clinical Prognostic Score for Death in Visceral Leishmaniasis Patients in a High HIV Co-Infection Burden Area in Ethiopia". In: *PLOS ONE* 12.6 (June 5, 2017). Ed. by Abhay R. Satoskar, e0178996. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0178996](https://doi.org/10.1371/journal.pone.0178996).
- [164] Dorcas Lamounier Costa et al. "Predicting Death from Kala-Azar: Construction, Development, and Validation of a Score Set and Accompanying Software". In: *Revista da Sociedade Brasileira de Medicina Tropical* 49.6 (Dec. 2016), pp. 728–740. ISSN: 0037-8682. DOI: [10.1590/0037-8682-0258-2016](https://doi.org/10.1590/0037-8682-0258-2016).
- [165] Costa, D. "Fatores de Prognóstico Na Leishmaniose Visceral: Alterações Clínicas e Laboratoriais Associadas à Resposta Imune, Aos Distúrbios Da Coagulação e à Morte". Universidade Federal de Minas Gerais, 2009. HDL: [1843/BU0S-96FFS7](https://hdl.handle.net/1843/BU0S-96FFS7).
- [166] Ministry of Health (Brazil). *Visceral Leishmaniasis: Clinical Recommendations for Lethality Reduction (Leishmaniose Visceral : Recomendações Clínicas Para Redução Da Letalidade)*. 1st ed. Brasilia, Brazil, 2011. ISBN: 978-85-334-1795-3.
- [167] Wendel Coura-Vital et al. "Prognostic Factors and Scoring System for Death from Visceral Leishmaniasis: An Historical Cohort Study in Brazil". In: *PLoS Neglected Tropical Diseases* 8.12 (Dec. 11, 2014). Ed. by Helton Da Costa Santiago, e3374. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0003374](https://doi.org/10.1371/journal.pntd.0003374).
- [168] G.L. Werneck et al. "Prognostic Factors for Death from Visceral Leishmaniasis in Teresina, Brazil". In: *Infection* 31.3 (June 2003), pp. 174–177. ISSN: 0300-8126, 1439-0973. DOI: [10.1007/s15010-003-3139-9](https://doi.org/10.1007/s15010-003-3139-9).
- [169] Gary S Collins et al. "External Validation of Multivariable Prediction Models: A Systematic Review of Methodological Conduct and Reporting". In: *BMC Medical Research Methodology* 14.1 (Dec. 2014), p. 40. ISSN: 1471-2288. DOI: [10.1186/1471-2288-14-40](https://doi.org/10.1186/1471-2288-14-40).
- [170] Kym I E Snell et al. "Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis: Checklist for Systematic Reviews and Meta-Analyses (TRIPOD-SRMA)". In: *BMJ* (May 3, 2023), e073538. ISSN: 1756-1833. DOI: [10.1136/bmj-2022-073538](https://doi.org/10.1136/bmj-2022-073538).
- [171] Karel G. M. Moons et al. "Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist". In: *PLoS Medicine* 11.10 (Oct. 14, 2014), e1001744. ISSN: 1549-1676. DOI: [10.1371/journal.pmed.1001744](https://doi.org/10.1371/journal.pmed.1001744).
- [172] Robert F. Wolff et al. "PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies". In: *Annals of Internal Medicine* 170.1 (Jan. 1, 2019), p. 51. ISSN: 0003-4819. DOI: [10.7326/M18-1376](https://doi.org/10.7326/M18-1376).
- [173] Karel G.M. Moons et al. "PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration". en. In: *Annals of Internal Medicine* 170.1 (Jan. 2019), W1–W33. ISSN: 0003-4819, 1539-3704. DOI: [10.7326/M18-1377](https://doi.org/10.7326/M18-1377).

- [174] Thomas P A Debray et al. “A guide to systematic review and meta-analysis of prediction model performance”. In: *BMJ* 356 (2017). DOI: [10.1136/bmj.i6460](https://doi.org/10.1136/bmj.i6460). eprint: <https://www.bmjjournals.org/content/356/bmj.i6460.full.pdf>.
- [175] Ewout W. Steyerberg et al. “Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research”. In: *PLoS Medicine* 10.2 (Feb. 5, 2013), e1001381. ISSN: 1549-1676. DOI: [10.1371/journal.pmed.1001381](https://doi.org/10.1371/journal.pmed.1001381).
- [176] Gary S. Collins et al. “Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement”. en. In: *Annals of Internal Medicine* 162.1 (Jan. 2015), pp. 55–63. ISSN: 0003-4819, 1539-3704. DOI: [10.7326/M14-0697](https://doi.org/10.7326/M14-0697).
- [177] Richard D Riley et al. “A Guide to Systematic Review and Meta-Analysis of Prognostic Factor Studies”. In: *BMJ* (Jan. 30, 2019), k4597. ISSN: 0959-8138, 1756-1833. DOI: [10.1136/bmj.k4597](https://doi.org/10.1136/bmj.k4597).
- [178] B. J. Ingui and M. A. M. Rogers. “Searching for Clinical Prediction Rules in MEDLINE”. In: *Journal of the American Medical Informatics Association* 8.4 (July 1, 2001), pp. 391–397. ISSN: 1067-5027, 1527-974X. DOI: [10.1136/jamia.2001.0080391](https://doi.org/10.1136/jamia.2001.0080391).
- [179] Geert-Jan Geersing et al. “Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews”. In: *PLoS ONE* 7.2 (Feb. 29, 2012). Ed. by Neil R. Smalheiser, e32844. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0032844](https://doi.org/10.1371/journal.pone.0032844).
- [180] Veritas Health Innovation. *Covidence systematic review software*. <https://www.covidence.org>. Melbourne, Australia.
- [181] Paul A. Harris et al. “The REDCap Consortium: Building an International Community of Software Platform Partners”. In: *Journal of Biomedical Informatics* 95 (July 2019), p. 103208. ISSN: 15320464. DOI: [10.1016/j.jbi.2019.103208](https://doi.org/10.1016/j.jbi.2019.103208).
- [182] Laura J Bonnett et al. “Guide to Presenting Clinical Prediction Models for Use in Clinical Settings”. In: *BMJ* (Apr. 17, 2019), p. l737. ISSN: 0959-8138, 1756-1833. DOI: [10.1136/bmj.l737](https://doi.org/10.1136/bmj.l737).
- [183] Valdelaine Etelvina Miranda De Araújo et al. “Early Clinical Manifestations Associated with Death from Visceral Leishmaniasis”. In: *PLoS Neglected Tropical Diseases* 6.2 (Feb. 7, 2012). Ed. by Richard Reithinger, e1511. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0001511](https://doi.org/10.1371/journal.pntd.0001511).
- [184] Márcia Jaqueline Alves De Queiroz Sampaio et al. “Risk Factors for Death in Children with Visceral Leishmaniasis”. In: *PLoS Neglected Tropical Diseases* 4.11 (Nov. 2, 2010). Ed. by Carlos Franco-Paredes, e877. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0000877](https://doi.org/10.1371/journal.pntd.0000877).
- [185] Suzette S. Kämink et al. “A Clinical Severity Scoring System for Visceral Leishmaniasis in Immunocompetent Patients in South Sudan”. In: *PLOS Neglected Tropical Diseases* 11.10 (Oct. 2, 2017). Ed. by Diana N. J. Lockwood, e0005921. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0005921](https://doi.org/10.1371/journal.pntd.0005921).
- [186] Juliana Foinquinos et al. “Temporal Validation of a Predictive Score for Death in Children with Visceral Leishmaniasis”. In: *Journal of Tropical Medicine* 2021 (Dec. 22, 2021). Ed. by Maoshui Wang, pp. 1–6. ISSN: 1687-9694, 1687-9686. DOI: [10.1155/2021/6688444](https://doi.org/10.1155/2021/6688444).

- [187] Benjamin S. Wessler et al. “Clinical Prediction Models for Cardiovascular Disease: Tufts Predictive Analytics and Comparative Effectiveness Clinical Prediction Model Database”. In: *Circulation. Cardiovascular Quality and Outcomes* 8.4 (July 2015), pp. 368–375. ISSN: 1941-7705. DOI: [10.1161/CIRCOUTCOMES.115.001693](https://doi.org/10.1161/CIRCOUTCOMES.115.001693). PMID: [26152680](#).
- [188] Iraci Duarte Lima et al. “Changing demographics of visceral leishmaniasis in northeast Brazil: Lessons for the future”. en. In: *PLOS Neglected Tropical Diseases* 12.3 (Mar. 2018). Ed. by David Harley, e0006164. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0006164](https://doi.org/10.1371/journal.pntd.0006164).
- [189] Ben Van Calster et al. “Calibration: the Achilles heel of predictive analytics”. en. In: *BMC Medicine* 17.1 (Dec. 2019), p. 230. ISSN: 1741-7015. DOI: [10.1186/s12916-019-1466-7](https://doi.org/10.1186/s12916-019-1466-7).
- [190] Walter Bouwmeester et al. “Reporting and Methods in Clinical Prediction Research: A Systematic Review”. In: *PLoS medicine* 9.5 (2012), pp. 1–12. ISSN: 1549-1676. DOI: [10.1371/journal.pmed.1001221](https://doi.org/10.1371/journal.pmed.1001221). PMID: [22629234](#).
- [191] Paula Dhiman et al. “Sample Size Requirements Are Not Being Considered in Studies Developing Prediction Models for Binary Outcomes: A Systematic Review”. In: *BMC Medical Research Methodology* 23.1 (Aug. 19, 2023), p. 188. ISSN: 1471-2288. DOI: [10.1186/s12874-023-02008-1](https://doi.org/10.1186/s12874-023-02008-1).
- [192] Ewout W. Steyerberg. “Overfitting and Optimism in Prediction Models”. en. In: *Clinical Prediction Models*. Cham: Springer International Publishing, 2019, pp. 95–112. ISBN: 9783030163983. DOI: [10.1007/978-3-030-16399-0_5](https://doi.org/10.1007/978-3-030-16399-0_5).
- [193] Orestis Efthimiou et al. “Developing clinical prediction models: a step-by-step guide”. en. In: *BMJ* 386 (2024), e078276. ISSN: 1756-1833. DOI: [10.1136/bmj-2023-078276](https://doi.org/10.1136/bmj-2023-078276).
- [194] Johanna A A G Damen et al. “Prediction models for cardiovascular disease risk in the general population: systematic review”. en. In: *BMJ* (May 2016), p. i2416. ISSN: 1756-1833. DOI: [10.1136/bmj.i2416](https://doi.org/10.1136/bmj.i2416).
- [195] Vanesa Bellou et al. “Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal”. en. In: *BMJ* (Oct. 2019), p. l5358. ISSN: 0959-8138, 1756-1833. DOI: [10.1136/bmj.l5358](https://doi.org/10.1136/bmj.l5358).
- [196] Gary S Collins et al. “Evaluation of clinical prediction models (part 1): from development to external validation”. en. In: *BMJ* (Jan. 2024), e074819. ISSN: 1756-1833. DOI: [10.1136/bmj-2023-074819](https://doi.org/10.1136/bmj-2023-074819).
- [197] Richard D Riley et al. “Evaluation of clinical prediction models (part 2): how to undertake an external validation study”. en. In: *BMJ* 384 (Jan. 2024), e074820. ISSN: 1756-1833. DOI: [10.1136/bmj-2023-074820](https://doi.org/10.1136/bmj-2023-074820).
- [198] Richard D. Riley et al. “Calculating the Sample Size Required for Developing a Clinical Prediction Model”. In: *BMJ (Clinical research ed.)* 368 (Mar. 18, 2020), p. m441. ISSN: 1756-1833. DOI: [10.1136/bmj.m441](https://doi.org/10.1136/bmj.m441). PMID: [32188600](#).
- [199] Richard D Riley et al. “Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study”. en. In: *BMJ* 384 (Jan. 2024), e074821. ISSN: 1756-1833. DOI: [10.1136/bmj-2023-074821](https://doi.org/10.1136/bmj-2023-074821).

- [200] Karel G M Moons et al. "PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods". en. In: *BMJ* 388 (Mar. 2025), e082505. ISSN: 1756-1833. DOI: [10.1136/bmj-2024-082505](https://doi.org/10.1136/bmj-2024-082505).
- [201] Charles Abongomera et al. "Prognostic factors for mortality among patients with visceral leishmaniasis in East Africa: Systematic review and meta-analysis". en. In: *PLOS Neglected Tropical Diseases* 14.5 (May 2020). Ed. by Guilherme L. Werneck, e0008319. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0008319](https://doi.org/10.1371/journal.pntd.0008319).
- [202] Vinícius Silva Belo et al. "Risk Factors for Adverse Prognosis and Death in American Visceral Leishmaniasis: A Meta-analysis". In: *PLoS Neglected Tropical Diseases* 8.7 (July 24, 2014). Ed. by Marleen Boelaert, e2982. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0002982](https://doi.org/10.1371/journal.pntd.0002982).
- [203] World Health Organization. *Regional strategic framework for accelerating and sustaining elimination of kala-azar in the South-East Asia Region, 2022-2026*. eng. OCLC: 1344416713. New Delhi: World Health Organization, South-East Asian Region, 2022.
- [204] Abhishek Kumar et al. "AI-driven analysis by identifying risk factors of VL relapse in HIV co-infected patients". In: *Scientific reports* 15.1 (July 2025), p. 21067. ISSN: 2045-2322. DOI: [10.1038/s41598-025-07406-7](https://doi.org/10.1038/s41598-025-07406-7).
- [205] André Cintas Donizette et al. "Predicting leishmaniasis outbreaks in Brazil using machine learning models based on disease surveillance and meteorological data". en. In: *Operations Research for Health Care* 44 (2025), p. 100453. ISSN: 22116923. DOI: [10.1016/j.orhc.2024.100453](https://doi.org/10.1016/j.orhc.2024.100453).
- [206] D. Campos. "[Predictive score of infectious complications on kalazar of child]". In: *Jornal De Pediatria* 72.6 (1996), pp. 411–414. ISSN: 0021-7557. PMID: [14688908](https://pubmed.ncbi.nlm.nih.gov/14688908/).
- [207] Tália S. Machado De Assis et al. "Predictive Models for the Diagnostic of Human Visceral Leishmaniasis in Brazil". In: *PLoS Neglected Tropical Diseases* 6.2 (Feb. 28, 2012). Ed. by Richard Reithinger, e1542. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0001542](https://doi.org/10.1371/journal.pntd.0001542).
- [208] Mohamed Slim Abdelmoula et al. "[Visceral leishmaniasis in children: prognostic factors]". In: *La Tunisie Medicale* 81.8 (Aug. 2003), pp. 535–539. ISSN: 0041-4131. PMID: [14608735](https://pubmed.ncbi.nlm.nih.gov/14608735/).
- [209] Thais Almeida Marques Da Silva et al. "Prognostic Factors Associated with Death from Visceral Leishmaniasis: A Case-Control Study in Brazil". In: *Transactions of The Royal Society of Tropical Medicine and Hygiene* 114.5 (May 7, 2020), pp. 346–354. ISSN: 0035-9203, 1878-3503. DOI: [10.1093/trstmh/traa010](https://doi.org/10.1093/trstmh/traa010).
- [210] Ingrid V. Oliveira-Sena and Guilherme L. Werneck. "Risk Factors for In-Hospital Mortality from Visceral Leishmaniasis: A Case-Control Study". In: *Journal of Infection and Public Health* 13.4 (Apr. 2020), pp. 538–543. ISSN: 18760341. DOI: [10.1016/j.jiph.2019.10.003](https://doi.org/10.1016/j.jiph.2019.10.003).
- [211] Institute of Tropical Medicine, Belgium. *Comprehensive Clinical and Laboratory Assessment of Visceral Leishmaniasis to Develop Clinical Prognostic Tools to Predict Relapse: a Prospective Cohort Study in Ethiopia*. ClinicalTrials.gov, NCT05602610. Prospective observational cohort study on prognostic tools for visceral leishmaniasis relapse in Ethiopia. 2024.

- [212] Thomas P A Debray et al. “Transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-Cluster): explanation and elaboration”. en. In: *BMJ* (Feb. 2023), e071058. ISSN: 1756-1833. DOI: [10.1136/bmj-2022-071058](https://doi.org/10.1136/bmj-2022-071058).
- [213] Maarten Van Smeden et al. “Clinical prediction models: diagnosis versus prognosis”. en. In: *Journal of Clinical Epidemiology* 132 (Apr. 2021), pp. 142–145. ISSN: 08954356. DOI: [10.1016/j.jclinepi.2021.01.009](https://doi.org/10.1016/j.jclinepi.2021.01.009).
- [214] Richard D. Riley et al. “IPD Meta-Analysis for Clinical Prediction Model Research”. In: *Individual Participant Data Meta-Analysis: A Handbook for Healthcare Research*. Ed. by Richard D. Riley et al. Hoboken, NJ: John Wiley & Sons, Inc., 2021, pp. 447–497. ISBN: 9781119333722. DOI: [10.1002/9781119333784.ch17](https://doi.org/10.1002/9781119333784.ch17).
- [215] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2025.
- [216] Jacob T. Bush et al. “Systematic review of clinical trials assessing the therapeutic efficacy of visceral leishmaniasis treatments: A first step to assess the feasibility of establishing an individual patient data sharing platform”. en. In: *PLOS Neglected Tropical Diseases* 11.9 (Sept. 2017). Ed. by Waleed Saleh Al-Salem, e0005781. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0005781](https://doi.org/10.1371/journal.pntd.0005781).
- [217] Sauman Singh-Phulgenda et al. “Infectious Diseases Data Observatory (IDDO) visceral leishmaniasis library of clinical therapeutic studies: A protocol for a living systematic review of clinical studies”. In: *Wellcome Open Research* 7 (2022). Study Protocol; published 18 May 2022, p. 155. DOI: [10.12688/wellcomeopenres.17739.1](https://doi.org/10.12688/wellcomeopenres.17739.1).
- [218] Infectious Diseases Data Observatory. *Data Governance*. <https://www.iddo.org/governance>. Accessed: 15 December 2025. 2025.
- [219] Prabin Dahal et al. “Addressing research questions in visceral leishmaniasis and post-kala-azar dermal leishmaniasis: Potential of reusing data from the Infectious Diseases Data Observatory platform”. en. In: *Annals of Medical Science & Research* 4.Suppl 1 (June 2025), S24–S34. ISSN: 2949-785X, 2949-7868. DOI: [10.4103/amr.amr_45_24](https://doi.org/10.4103/amr.amr_45_24).
- [220] Clinical Data Interchange Standards Consortium (CDISC). *Study Data Tabulation Model (SDTM)*. Tech. rep.
- [221] Infectious Diseases Data Observatory. *New global collaboration developing standards for standardised data collection across the VL research community*. 2020. URL: <https://www.iddo.org/news/new-global-collaboration-developing-standards-standardised-data-collection-across-vl-research> (visited on 10/06/2025).
- [222] Hadley Wickham et al. “Welcome to the tidyverse”. In: *Journal of Open Source Software* 4.43 (2019), p. 1686. DOI: [10.21105/joss.01686](https://doi.org/10.21105/joss.01686).
- [223] Infectious Diseases Data Observatory (IDDO). *Visceral leishmaniasis data inventory*. <https://www.iddo.org/document/visceral-leishmaniasis-data-inventory>. Last updated November 2025; accessed December 18, 2025. Nov. 2025.

- [224] Vassiliki Syriopoulou et al. “Two Doses of a Lipid Formulation of Amphotericin B for the Treatment of Mediterranean Visceral Leishmaniasis”. en. In: *Clinical Infectious Diseases* 36.5 (Mar. 2003), pp. 560–566. ISSN: 1058-4838, 1537-6591. DOI: [10.1086/367843](https://doi.org/10.1086/367843).
- [225] Gustavo Adolfo Sierra Romero et al. “Efficacy and safety of available treatments for visceral leishmaniasis in Brazil: A multicenter, randomized, open label trial”. en. In: *PLOS Neglected Tropical Diseases* 11.6 (2017). Ed. by Richard Reithinger, e0005706. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0005706](https://doi.org/10.1371/journal.pntd.0005706).
- [226] Peter Peduzzi et al. “A simulation study of the number of events per variable in logistic regression analysis”. In: *Journal of Clinical Epidemiology* 49.12 (Dec. 1996), pp. 1373–1379. ISSN: 0895-4356. DOI: [10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3).
- [227] Frank E. Harrell. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. en. Springer Series in Statistics. Cham: Springer International Publishing, 2015. ISBN: 9783319194240. DOI: [10.1007/978-3-319-19425-7](https://doi.org/10.1007/978-3-319-19425-7).
- [228] Delphine S. Courvoisier et al. “Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure”. en. In: *Journal of Clinical Epidemiology* 64.9 (2011), pp. 993–1000. ISSN: 08954356. DOI: [10.1016/j.jclinepi.2010.11.012](https://doi.org/10.1016/j.jclinepi.2010.11.012).
- [229] Richard D Riley et al. “Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes”. en. In: *Statistics in Medicine* 38.7 (Mar. 2019), pp. 1276–1296. ISSN: 0277-6715, 1097-0258. DOI: [10.1002/sim.7992](https://doi.org/10.1002/sim.7992).
- [230] Joie Ensor. *pmsampsize: Sample Size for Development of a Prediction Model*. 2023.
- [231] World Health Organization. *WHO guideline on the prevention and management of wasting and nutritional oedema (acute malnutrition) in infants and children under 5 years*. Accessed: 2025-12-28. 2023. ISBN: 978-92-4-008283-0.
- [232] Mercedes De Onis. “Development of a WHO growth reference for school-aged children and adolescents”. In: *Bulletin of the World Health Organization* 85.09 (2007), pp. 660–667. ISSN: 00429686. DOI: [10.2471/BLT.07.043497](https://doi.org/10.2471/BLT.07.043497).
- [233] Tim J Cole et al. “Body mass index cut offs to define thinness in children and adolescents: international survey”. en. In: *BMJ* 335.7612 (2007), p. 194. ISSN: 0959-8138, 1468-5833. DOI: [10.1136/bmj.39238.399444.55](https://doi.org/10.1136/bmj.39238.399444.55).
- [234] Dirk Schumacher. *anthro: Computation of the WHO Child Growth Standards*. 2023.
- [235] Dirk Schumacher. *anthroplus: Computation of the WHO 2007 References for School-age Children and Adolescents (5 to 19 Years)*. 2021.
- [236] World Health Organization, ed. *Guideline on haemoglobin cutoffs to define anaemia in individuals and populations*. eng. Geneva: World Health Organization, 2024. ISBN: 978-92-4-008854-2.
- [237] Barret Schloerke et al. *GGally: Extension to 'ggplot2'*. R package version 2.4.0. 2025.

- [238] M J Gardner and D G Altman. “Confidence intervals rather than P values: estimation rather than hypothesis testing.” en. In: *BMJ* 292.6522 (Mar. 1986), pp. 746–750. ISSN: 0267-0623, 1468-5833. DOI: [10.1136/bmj.292.6522.746](https://doi.org/10.1136/bmj.292.6522.746).
- [239] S. N. Wood. “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models”. In: *Journal of the Royal Statistical Society (B)* 73.1 (2011), pp. 3–36. DOI: [10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).
- [240] Edwin B. Wilson. “Probable Inference, the Law of Succession, and Statistical Inference”. en. In: *Journal of the American Statistical Association* 22.158 (1927), pp. 209–212. ISSN: 0162-1459, 1537-274X. DOI: [10.1080/01621459.1927.10502953](https://doi.org/10.1080/01621459.1927.10502953).
- [241] Ian R. White et al. “Multiple imputation using chained equations: Issues and guidance for practice”. en. In: *Statistics in Medicine* 30.4 (Feb. 2011), pp. 377–399. ISSN: 0277-6715, 1097-0258. DOI: [10.1002/sim.4067](https://doi.org/10.1002/sim.4067).
- [242] Stef van Buuren. *Flexible imputation of missing data*. eng. Second edition. Chapman & Hall/CRC interdisciplinary statistics series. Boca Raton: Chapman & Hall/CRC, 2021. ISBN: 978-1-03-217863-9.
- [243] Stef van Buuren and Karin Groothuis-Oudshoorn. “mice: Multivariate Imputation by Chained Equations in R”. In: *Journal of Statistical Software* 45.3 (2011), pp. 1–67. DOI: [10.18637/jss.v045.i03](https://doi.org/10.18637/jss.v045.i03).
- [244] Vincent Audigier and Matthieu Resche-Rigon. *micemd: Multiple Imputation by Chained Equations with Multilevel Data*. 2023.
- [245] Donald B. Rubin. *Multiple imputation for nonresponse in surveys*. eng. Wiley series in probability and mathematical statistics: Applied probability and statistics. New York: Wiley, 1987. DOI: [10.1002/9780470316696](https://doi.org/10.1002/9780470316696).
- [246] Vincent Audigier et al. “Multiple Imputation for Multilevel Data with Continuous and Binary Variables”. In: *Statistical Science* 33.2 (May 2018). ISSN: 0883-4237. DOI: [10.1214/18-STS646](https://doi.org/10.1214/18-STS646).
- [247] Kristian Kleinke. *countimp: Multiple Imputation of incomplete count data*. 2024.
- [248] Walter Bouwmeester et al. “Prediction models for clustered data: comparison of a random intercept and standard regression model”. en. In: *BMC Medical Research Methodology* 13.1 (Dec. 2013), p. 19. ISSN: 1471-2288. DOI: [10.1186/1471-2288-13-19](https://doi.org/10.1186/1471-2288-13-19).
- [249] L Wynants et al. “Does ignoring clustering in multicenter data influence the performance of prediction models? A simulation study”. en. In: *Statistical Methods in Medical Research* 27.6 (June 2018), pp. 1723–1736. ISSN: 0962-2802, 1477-0334. DOI: [10.1177/0962280216668555](https://doi.org/10.1177/0962280216668555).
- [250] Ewout W. Steyerberg et al. “Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: An overview and illustration”. eng. In: *Statistics in Medicine* 38.22 (2019), pp. 4290–4309. DOI: [10.1002/sim.8296](https://doi.org/10.1002/sim.8296).
- [251] Douglas Bates et al. “Fitting Linear Mixed-Effects Models Using **lme4**”. en. In: *Journal of Statistical Software* 67.1 (2015). ISSN: 1548-7660. DOI: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).

- [252] Peter C. Austin et al. “Effect of Variable Selection Strategy on the Performance of Prognostic Models When Using Multiple Imputation”. en. In: *Circulation: Cardiovascular Quality and Outcomes* 12.11 (Nov. 2019), e005927. ISSN: 1941-7713, 1941-7705. DOI: [10.1161/CIRCOUTCOMES.119.005927](https://doi.org/10.1161/CIRCOUTCOMES.119.005927).
- [253] Angela M. Wood et al. “How should variable selection be performed with multiply imputed data?” en. In: *Statistics in Medicine* 27.17 (2008), pp. 3227–3246. ISSN: 0277-6715, 1097-0258. DOI: [10.1002/sim.3177](https://doi.org/10.1002/sim.3177).
- [254] K. H. Li et al. “Large-Sample Significance Levels from Multiply Imputed Data Using Moment-Based Statistics and an F Reference Distribution”. In: *Journal of the American Statistical Association* 86.416 (Dec. 1991), p. 1065. ISSN: 01621459. DOI: [10.2307/2290525](https://doi.org/10.2307/2290525).
- [255] David Van Klaveren et al. “Assessing discriminative ability of risk models in clustered data”. en. In: *BMC Medical Research Methodology* 14.1 (Dec. 2014), p. 5. ISSN: 1471-2288. DOI: [10.1186/1471-2288-14-5](https://doi.org/10.1186/1471-2288-14-5).
- [256] Xavier Robin et al. “pROC: an open-source package for R and S+ to analyze and compare ROC curves”. In: *BMC Bioinformatics* 12 (2011), p. 77.
- [257] Stephen Burgess et al. “Combining multiple imputation and meta-analysis with individual participant data”. eng. In: *Statistics in Medicine* 32.26 (Nov. 2013), pp. 4499–4514. ISSN: 1097-0258. DOI: [10.1002/sim.5844](https://doi.org/10.1002/sim.5844).
- [258] Guido Knapp and Joachim Hartung. “Improved tests for a random effects meta-regression with a single covariate”. en. In: *Statistics in Medicine* 22.17 (2003), pp. 2693–2710. ISSN: 0277-6715, 1097-0258. DOI: [10.1002/sim.1482](https://doi.org/10.1002/sim.1482).
- [259] R. Van Oirbeek and E. Lesaffre. “Assessing the predictive ability of a multilevel binary regression model”. en. In: *Computational Statistics & Data Analysis* 56.6 (2012), pp. 1966–1980. ISSN: 01679473. DOI: [10.1016/j.csda.2011.11.023](https://doi.org/10.1016/j.csda.2011.11.023).
- [260] Nancy A. Obuchowski and Michael L. Lieber. “Confidence intervals for the receiver operating characteristic area in studies with small samples”. en. In: *Academic Radiology* 5.8 (Aug. 1998), pp. 561–571. ISSN: 10766332. DOI: [10.1016/S1076-6332\(98\)80208-0](https://doi.org/10.1016/S1076-6332(98)80208-0).
- [261] Blaise Hanczar et al. “Small-sample precision of ROC-related estimates”. en. In: *Bioinformatics* 26.6 (Mar. 2010), pp. 822–830. ISSN: 1367-4811, 1367-4803. DOI: [10.1093/bioinformatics/btq037](https://doi.org/10.1093/bioinformatics/btq037).
- [262] Paul-Christian Bürkner. “brms: An R Package for Bayesian Multilevel Models Using Stan”. en. In: *Journal of Statistical Software* 80.1 (2017). ISSN: 1548-7660. DOI: [10.18637/jss.v080.i01](https://doi.org/10.18637/jss.v080.i01).
- [263] Stan Development Team. *Stan Reference Manual*. Version 2.37. Stan. 2025. URL: <https://mc-stan.org/docs/reference-manual/>.
- [264] W. Bouwmeester et al. “Internal Validation of Risk Models in Clustered Data: A Comparison of Bootstrap Schemes”. en. In: *American Journal of Epidemiology* 177.11 (June 2013), pp. 1209–1217. ISSN: 1476-6256, 0002-9262. DOI: [10.1093/aje/kws396](https://doi.org/10.1093/aje/kws396).

- [265] D. Chakraborty et al. "Human placental extract offers protection against experimental visceral leishmaniasis: a pilot study for a phase-I clinical trial". en. In: *Annals of Tropical Medicine & Parasitology* 102.1 (Jan. 2008), pp. 21–38. ISSN: 0003-4983, 1364-8594. DOI: [10.1179/136485908X252133](https://doi.org/10.1179/136485908X252133).
- [266] Vidya Nand Rabi Das et al. "A controlled, randomized nonblinded clinical trial to assess the efficacy of amphotericin B deoxycholate as compared to pentamidine for the treatment of antimony unresponsive visceral leishmaniasis cases in Bihar, India". eng. In: *Therapeutics and Clinical Risk Management* 5.1 (Feb. 2009), pp. 117–124. ISSN: 1176-6336.
- [267] S. Koirala. "Phase IV Trial of Miltefosine in the Treatment of Visceral Leishmaniasis". Protocol only; no results published. Protocol not publicly available. Open-label Phase IV safety and efficacy study of oral miltefosine (28 days), dosed according to age and weight. Dharan, Nepal, 2003.
- [268] Krishna Pandey et al. "Pharmacovigilance of Miltefosine in Treatment of Visceral Leishmaniasis in Endemic Areas of Bihar, India". In: *The American Society of Tropical Medicine and Hygiene* 95.5 (Nov. 2016), pp. 1100–1105. ISSN: 0002-9637, 1476-1645. DOI: [10.4269/ajtmh.16-0242](https://doi.org/10.4269/ajtmh.16-0242).
- [269] Krishna Pandey et al. "Efficacy and Safety of Liposomal Amphotericin B for Visceral Leishmaniasis in Children and Adolescents at a Tertiary Care Center in Bihar, India". In: *The American Journal of Tropical Medicine and Hygiene* 97.5 (Nov. 2017), pp. 1498–1502. ISSN: 0002-9637, 1476-1645. DOI: [10.4269/ajtmh.17-0094](https://doi.org/10.4269/ajtmh.17-0094).
- [270] S. Rijal et al. "Treatment of visceral leishmaniasis in south-eastern Nepal: decreasing efficacy of sodium stibogluconate and need for a policy to limit further decline". en. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 97.3 (May 2003), pp. 350–354. ISSN: 00359203. DOI: [10.1016/S0035-9203\(03\)90167-2](https://doi.org/10.1016/S0035-9203(03)90167-2).
- [271] Suman Rijal. *Efficacy and safety of liposomal amphotericin B in Nepalese patients with visceral leishmaniasis*. Australian New Zealand Clinical Trials Registry ACTRN12610000130066. Accessed 17 June 2025. 2010.
- [272] S. Rijal et al. "Clinical risk factors for therapeutic failure in kala-azar patients treated with pentavalent antimonials in Nepal". en. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 104.3 (Mar. 2010), pp. 225–229. ISSN: 00359203. DOI: [10.1016/j.trstmh.2009.08.002](https://doi.org/10.1016/j.trstmh.2009.08.002).
- [273] Shyam Sundar et al. "New Treatment Approach in Indian Visceral Leishmaniasis: Single-Dose Liposomal Amphotericin B Followed by Short-Course Oral Miltefosine". en. In: *Clinical Infectious Diseases* 47.8 (Oct. 2008), pp. 1000–1006. ISSN: 1058-4838, 1537-6591. DOI: [10.1086/591972](https://doi.org/10.1086/591972).
- [274] S. Sundar et al. "Safety of a pre-formulated amphotericin B lipid emulsion for the treatment of Indian Kala-azar". en. In: *Tropical Medicine & International Health* 13.9 (Sept. 2008), pp. 1208–1212. ISSN: 1360-2276, 1365-3156. DOI: [10.1111/j.1365-3156.2008.02128.x](https://doi.org/10.1111/j.1365-3156.2008.02128.x).

- [275] Shyam Sundar et al. "Short-Course Paromomycin Treatment of Visceral Leishmaniasis in India: 14-Day vs 21-Day Treatment". en. In: *Clinical Infectious Diseases* 49.6 (Sept. 2009), pp. 914–918. ISSN: 1058-4838, 1537-6591. DOI: [10.1086/605438](https://doi.org/10.1086/605438).
- [276] Shyam Sundar et al. "Single-Dose Liposomal Amphotericin B for Visceral Leishmaniasis in India". en. In: *New England Journal of Medicine* 362.6 (Feb. 2010), pp. 504–512. ISSN: 0028-4793, 1533-4406. DOI: [10.1056/NEJMoa0903627](https://doi.org/10.1056/NEJMoa0903627).
- [277] Shyam Sundar et al. "Ambisome plus miltefosine for Indian patients with kala-azar". en. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 105.2 (Feb. 2011), pp. 115–117. ISSN: 00359203. DOI: [10.1016/j.trstmh.2010.10.008](https://doi.org/10.1016/j.trstmh.2010.10.008).
- [278] Shyam Sundar et al. "Efficacy and Safety of Amphotericin B Emulsion versus Liposomal Formulation in Indian Patients with Visceral Leishmaniasis: A Randomized, Open-Label Study". en. In: *PLoS Neglected Tropical Diseases* 8.9 (Sept. 2014). Ed. by Elodie Ghedin, e3169. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0003169](https://doi.org/10.1371/journal.pntd.0003169).
- [279] Shyam Sundar et al. "Single-Dose Indigenous Liposomal Amphotericin B in the Treatment of Indian Visceral Leishmaniasis: A Phase 2 Study". In: *The American Society of Tropical Medicine and Hygiene* 92.3 (Mar. 2015), pp. 513–517. ISSN: 0002-9637, 1476-1645. DOI: [10.4269/ajtmh.14-0259](https://doi.org/10.4269/ajtmh.14-0259).
- [280] Shyam Sundar. *Safety and efficacy study of paromomycin to treat visceral leishmaniasis*. ClinicalTrials.gov (NCT00216346). Accessed 19 June 2025. 2005.
- [281] Shyam Sundar. *Combination Chemotherapy for the Treatment of Indian Visceral Leishmaniasis: Miltefosine Plus Liposomal Amphotericin B - Dose and Duration Ranging Study*. ClinicalTrials.gov (NCT00370825). Accessed 17 June 2025. 2008.
- [282] Shyam Sundar. *An Open Label Randomised Study to Assess the Safety and Efficacy of Short Course Paromomycin in Visceral Leishmaniasis*. ClinicalTrials.gov (NCT00629031). Accessed 18 June 2025. 2008.
- [283] Shyam Sundar. *Single Infusion of Liposomal Amphotericin B in Indian Visceral Leishmaniasis*. ClinicalTrials.gov (NCT00628719). Accessed 19 June 20025. 2008.
- [284] Shyam Sundar. *Short Course of Miltefosine and Liposomal Amphotericin B for Kala-azar*. ClinicalTrials.gov (NCT00371995). Accessed June 18 2025. 2006.
- [285] Shyam Sundar. *To Study the Effect Of Single Infusions Of Amphotericin B Lipid Preparations in Treatment of Patients Of Kala Azar*. ClinicalTrials.gov (NCT00876824). Access 18 June 2025. 2009.
- [286] Shyam Sundar. *A prospective, open-label, non-comparative, sequential, phase II, multicenter study to assess safety of single dose regimen at two dose levels of FUNCISOME TM in the treatment of visceral leishmaniasis (kala-azar)*. Clinical Trials Registry - India (CTRI/2011/11/002145). Accessed 19 June 2025. 2011.
- [287] Shyam Sundar. *Single dose liposomal amphotericin B for visceral leishmaniasis*. ClinicalTrials.gov (NCT01566552). Accessed 19 June 2025. 2012.

- [288] K. Ritmeijer et al. “Ethiopian visceral leishmaniasis: generic and proprietary sodium stibogluconate are equivalent; HIV co-infected patients have a poor outcome.” eng. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 95.6 (Dec. 2001), pp. 668–672. ISSN: 0035-9203. DOI: [10.1016/s0035-9203\(01\)90110-5](https://doi.org/10.1016/s0035-9203(01)90110-5).
- [289] Koert Ritmeijer et al. “A comparison of miltefosine and sodium stibogluconate for treatment of visceral leishmaniasis in an Ethiopian population with high prevalence of HIV infection.” eng. In: *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 43.3 (Aug. 2006), pp. 357–364. ISSN: 1537-6591 1058-4838. DOI: [10.1086/505217](https://doi.org/10.1086/505217).
- [290] H. Veeken et al. “A randomized comparison of branded sodium stibogluconate and generic sodium stibogluconate for the treatment of visceral leishmaniasis under field conditions in Sudan.” eng. In: *Tropical medicine & international health: TM & IH* 5.5 (May 2000), pp. 312–317. ISSN: 1360-2276. DOI: [10.1046/j.1365-3156.2000.00555.x](https://doi.org/10.1046/j.1365-3156.2000.00555.x).
- [291] Monique Wasunna et al. “Efficacy and Safety of AmBisome in Combination with Sodium Stibogluconate or Miltefosine and Miltefosine Monotherapy for African Visceral Leishmaniasis: Phase II Randomized Trial”. en. In: *PLOS Neglected Tropical Diseases* 10.9 (2016). Ed. by Philippe Büscher, e0004880. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0004880](https://doi.org/10.1371/journal.pntd.0004880).