

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Cohort Profile: ReCoDID Consortium’s Harmonized Acute Febrile Illness Arbovirus

Meta-Cohort

ReCoDID Pooled Cohort Profile

Gustavo Gómez^{1°}, Heather Hufstedler^{2°}, Carlos Montenegro^{3°}, Yannik Roell⁴, Anyela Lozano¹,
Adriana Tami^{5,6}, Tereza Magalhaes^{7, 8}, Ernesto T.A. Marques^{9,10}, Angel Balmaseda^{3,11},
Guilherme Calvet¹², Eva Harris¹³, Patricia Brasil¹², Victor Herrera¹, Luis Villar^{1,14}, Lauren
Maxwell², Thomas Jaenisch^{2, 4, 15}, ReCoDID Arbovirus harmonization study group*

1. *Grupo de Epidemiología Clínica, Universidad Industrial de Santander, Bucaramanga, Colombia*
2. *Heidelberg Institute of Global Health (HIGH), Heidelberg University Hospital, Heidelberg, Germany*
3. *Sustainable Sciences Institute, Managua, Nicaragua*
4. *Center for Global Health, Colorado School of Public Health, Aurora, Colorado, United States of America*
5. *University of Groningen, University Medical Center Groningen, Department of Medical Microbiology and Infection Prevention, Groningen, The Netherlands*
6. *Departamento de Estudios Clínicos, Facultad de Ciencias de la Salud, Universidad de Carabobo, Valencia, Venezuela*
7. *Center for Vector-Borne Infectious Diseases (CVID), Department of Microbiology, Immunology and Pathology (MIP), College of Veterinary Medicine and Biomedical Sciences, Fort Collins, CO, USA*

- 25 8. *Department of Preventive and Social Medicine, School of Medicine, Universidade*
26 *Federal da Bahia, Salvador, BA, Brazil*
- 27 9. *Institute Aggeu Magalhães, Department of Virology and Experimental Therapeutics,*
28 *Fundação Oswaldo Cruz, FIOCRUZ-PE, Recife Brazil*
- 29 10. *University of Pittsburgh, School of Public Health, Department of Infectious Diseases and*
30 *Microbiology, Pittsburgh, PA, USA*
- 31 11. *Laboratorio Nacional de Virología, Centro Nacional de Diagnóstico y Referencia,*
32 *Ministry of Health, Managua, Nicaragua.*
- 33 12. *Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brasil*
- 34 13. *Division of Infectious Diseases, School of Public Health, University of California*
35 *Berkeley, Berkeley, CA, USA*
- 36 14. *Centro de Atención y Diagnóstico de Enfermedades Infecciosas, Bucaramanga,*
37 *Colombia*
- 38 15. *Section Clinical Tropical Medicine, Department for Infectious Diseases, Heidelberg*
39 *University Hospital, Heidelberg, Germany*
- 40 ° *Co-first authors, in alphabetical order*
- 41 * *Membership of the ReCoDID Arbovirus harmonization study group is provided in the*
42 *acknowledgements*

43 Abstract

44 Infectious disease (ID) cohorts are key to advancing public health surveillance, public policies and
45 pandemic responses. Unfortunately, ID cohorts often lack funding to store and share clinical-
46 epidemiological data (CE) and high-dimensional laboratory (HDL) data long-term, which is
47 evident when the link between these data elements is not kept up to date. This becomes particularly
48 apparent when smaller cohorts fail to successfully address the initial scientific objectives due to
49 limited case numbers, which also limits the potential of pooling for these studies to monitor long-
50 term cross-disease interactions within and across populations. To facilitate advancements in cross-
51 population inference and reuse of cohort data, the European Commission (EC) and the Canadian
52 Institutes of Health Research, Institute of Genetics (CIHR-IG) funded the ReCoDID
53 (Reconciliation of Cohort Data for Infectious Diseases) Consortium to store and share harmonized
54 and standardized CE and HDL data on a federated platform and also provide innovative statistical
55 tools to conduct meta-analyses of the individual patient data. Here we describe the harmonization
56 of CE data from nine arbovirus (arthropod-borne viruses) cohorts in Latin America, which serve
57 as a starting point for the ReCoDID meta-cohort. CE data was retrospectively harmonized using
58 Maelstrom's methodology and standardized to Clinical Data Interchange Standards Consortium
59 (CDISC).

60 This meta-cohort will facilitate various joint research projects, e.g., on immunological interactions
61 between sequential flavivirus infections and for the evaluation of potential biomarkers for severe
62 arboviral disease.

63

64

65

66

67 Introduction

68 The meta-cohort consists of participant-level data and descriptive metadata from nine studies
69 from five countries (Brazil, Colombia, El Salvador, Nicaragua, and Venezuela). All studies were
70 established to study arbovirus (arthropod-borne viruses) infections in the population, with some
71 cohorts enrolling maternal-infant pairs, or either children or pregnant women.

72 Most studies did not recruit participants through study-initiated contacts (e.g., emails, calls, letters,
73 etc.). Instead, the vast majority of participants in each cohort were directly referred to the study
74 from the health unit where they were seeking care. A small fraction of participants contacted the
75 cohorts directly to participate because participation was associated with access to more routine
76 care or additional screenings, which can be seen as a benefit in resource-poor settings.

77 The interactions between immune responses caused by different patterns of exposure over time to
78 the four dengue virus serotypes (DENV 1-4) and Zika virus (ZIKV) have attracted considerable
79 attention - for example as a mechanism to explain the heterogeneity in severe dengue but also in
80 the severe outcomes seen during and after the ZIKV epidemic in Latin America(1–5). The
81 investigation of these interactions between closely related members of the *flaviviridae* requires
82 large sample sizes and the inclusion of populations with exposures to different sequences of
83 pathogens, resulting in heterogeneous immune profiles.

84

85 The ReCoDID Consortium, funded by the EC and CIHR-IG(6), aims to provide ID researchers
86 with harmonized participant-level data and metadata resources as well as analysis tools to facilitate
87 pooled analysis projects, i.e. to advance our knowledge on the effects of prior exposure on the
88 immune response to subsequent epidemics at the population and individual level and to inform
89 personalized medicine approaches to diagnosis and treatment of infections. To facilitate cross-

90 study inference in the context of emerging IDs, ReCoDID researchers created a platform to extract
91 individual-level CE and HDL data from existing cohorts, and harmonize this data according to a
92 specific standard. ReCoDID focuses current harmonization efforts on arbovirus and SARS-CoV-
93 2 cohort data, but hopes to expand these services to other IDs in the future. This cohort profile
94 provides an overview of the newly-created arbovirus meta-cohort from five countries: Brazil,
95 Colombia, El Salvador, Nicaragua, and Venezuela. Acute and post-acute samples were collected
96 from each study. Information extracted from samples vary from study to study and include DENV
97 molecular tests, DENV serotype, DENV viral load, ZIKV molecular test,, ZIKV viral load,
98 CHIKV molecular tests and CHIKV viral load. Height, weight, birthdate, negative health
99 outcomes associated with severe dengue (such as occurrence of bleeding, e.g.) and required
100 interventions have also been collected. The ReCoDID Consortium aims to build a data sharing
101 platform for linking CE data to HDL data (e.g., human and pathogen genomic data, human
102 metabolomic and immunomics data) at the participant level that are collected from ID focused
103 cohorts. While ReCoDID is working to share data related to other disease types, this paper
104 describes the acute febrile illness (AFI) meta-cohort that includes, at the date of publishing this
105 article, data from nine studies that have committed to sharing CE and HDL data from their
106 arbovirus cohorts. All participating cohorts have submitted genomic sequences of the DENV virus
107 to ReCoDID, except for the cohorts in Nicaragua and the Cohort of Symptomatic Pregnant Women
108 in Brazil. Two studies— PHBDC and IDAMS cohorts— have also agreed to share genomic
109 sequences for CHIKV and ZIKV viruses. Most participating cohorts collected and stored blood
110 samples; Nicaragua also collected urine and saliva samples. Cohorts varied in their inclusion
111 criteria— some admitting only patients who present with fever, some used rash or red eyes, while
112 others admitted patients who presented with fever and/or rash. With the introduction of ZIKV and

113 CHIKV, the PHBDC study chose to adjust the inclusion criteria in order to admit patients who
 114 presented with rash, headache and/or fever. Altogether, the longitudinal data of the meta-cohort
 115 covers over 18,000 patients (pediatric and adults), in both inpatient and outpatient settings from
 116 five countries (Brazil, Colombia, El Salvador, Nicaragua, and Venezuela). Data collection start
 117 and end dates vary between cohorts, but ranged from 1998 to the present where the patients have
 118 been followed up during different intervals as it can be identified in the table 1 below:

119

120 Table 1. Frequency of follow-up across included cohort studies

Study	Arm	Frequency of clinical observations	Frequency of laboratory investigations (not for diagnosis)	Frequency of laboratory Diagnosis (example 0,7,14)	
Col: Diagnosis Cohort: Evaluation of the diagnostic accuracy and usefulness of rapid tests for early diagnosis of dengue	Outpatient/Inpatient	3-5, 4-6, 5-7, 15-17	3-5, 4-6, 5-7, 15-17	3-5, 4-6, 5-7, 15-17	
Col: Prognosis: Immune Mechanisms of Pathogenesis in Patients with Dengue Infection	Outpatient/Inpatient	Once per day (Inpatient)	0-7 days, 3-6, and 24 weeks Note: Inpatients also provided samples every 48 hours during hospitalization (up to 4 samples)	0	
Col: Piedecuesta's household-based dynamic cohort. Identification of age groups to be prioritized for vaccination in a population of children and adolescents	Outpatient	Biweekly (All, telephone) Once per day (Incident Febrile cases)	Once per day (Incident Febrile Cases)	Annual (All) 0, 7 days (Incident Febrile cases)	
Col:AEDES Cohorts	Identification of prognostic markers of severity in dengue	Outpatient	Once per day	Once per day	0,14
	Validation of a clinical definition for dengue and evaluation of its usefulness to identify early conditions associated with hospitalization	Outpatient	Once per day	Once per day	0,14
	"AEDES: Abordando Áreas Endémicas del Dengue para el Estudio de su Severidad". Colombian multicentric study.	Outpatient	Once per day	0, 6	0,10
		Inpatient	Once per day	Once per day	0,10
CDY15: Efficacy and safety of a new tetravalent dengue vaccine in healthy children and adolescents aged 9 to 16 years in Latin America	Placebo	Enrollment, 6 y 12 (Months) (Incident Febrile Cases)	Daily (Incident Febrile Cases)	0-5, 14 (Incident Febrile Cases)	
Nica-Pediatric Dengue Hospital-based Study -	Outpatient	Once per day	Once per day	0,1,2;14-28	
	Inpatient	More than once per day	Once per day	0,1,2,14-28	
Nica-PDCS	Follow-up	2-3 times per week	2-3 times per week	0,14-21	
	Annual sample		Once per year.	0	

IDAMS	Recruitment as outpatients, some proceeded to hospitalization	Once per day	Once per day	0,3-6,15 (days)
Cohort of symptomatic pregnant women		At enrollment, weekly telephone follow-up, and a second visit within 30 days after enrollment	NA	0, 30

121

122

123 *Evaluation of the diagnostic accuracy and usefulness of rapid tests for early diagnosis of dengue*

124 *(Diagnosis Cohort)*

125 The diagnosis cohort, funded by E25bio, Inc., was aimed to determine the diagnostic usefulness
 126 of repeated NS1 rapid testing in clinical settings(7). This cohort enrolled and followed patients (≥ 2
 127 years old) who had both clinically suspected dengue and a positive dengue rapid test (NS1 antigen)
 128 at the time of consultation or hospitalization. Participants were followed at 1, 2, and 10 days
 129 (convalescence) after recruitment to determine the incidence of dengue complications among
 130 confirmed cases. Dengue infection was defined as positive NS1 results (acute sample).

131

132

133 *Immune Mechanisms of Pathogenesis in Patients with Dengue Infection (Prognosis Cohort)*

134 The Prognosis Cohort began with the goal to prospectively validate the predictive accuracy of a
 135 pool of transcriptomics intended to predict severity among patients with confirmed DENV
 136 infection(8). This study was funded by the U.S. Department of Defense and Colombia’s Centro de
 137 Atención y Diagnóstico de Enfermedades Infecciosas (CDI, Colombia). This cohort enrolled
 138 patients (≥ 1 year old) with clinically suspected dengue and conducted follow-up at 1, 2, and 10
 139 days (convalescence) after recruitment to determine the incidence of dengue complications among
 140 confirmed cases. Participants recruited in outpatient settings were clinically evaluated at enrolment

141 and asked to provide additional blood samples between 3-6 days and 7-10 days if their first sample
142 was obtained up to 2 days and 3-4 days after the onset of fever, respectively. Those who were
143 recruited from inpatient settings underwent daily clinical evaluation, and blood samples were
144 drawn every 48 hours during hospitalization (up to 4 samples). In all participants, additional
145 samples were collected 3-6 and 24 weeks after onset of fever(9,10). Dengue infection was defined
146 as positive PCR results (acute sample).

147

148 *Piedecuesta's Household-Based Dynamic Cohort (PHBDC)*

149 Piedecuesta's household-based dynamic cohort sought to estimate age-specific dengue
150 seroprevalence and identify age groups to be prioritized for vaccination among children and
151 adolescents. This study was funded by the Colombian Science Ministry, Minciencias. PHBDC was
152 a population-based cross-sectional study which began in 2014 and enrolled and evaluated healthy
153 children and adults (15%) from Piedecuesta (a mid-size city with endemic DENV). The aim of the
154 study was to estimate age-specific dengue seroprevalence in order to identify age groups to be
155 prioritized for vaccination among children and adolescents. Based on the results of this
156 seroprevalence study, a household-based dynamic cohort was initiated in 2015 with the aim of
157 estimating the age-specific incidence of dengue in Piedecuesta (N=2,730). This cohort enrolled
158 children (2-15 years old) and adults within the same household. The cohort followed up with
159 participants on a biweekly basis through telephone contact to identify incident febrile cases. Cases
160 were identified through clinical evaluation, and blood samples studied using Luminex ArboMIA
161 to determine etiological diagnosis (annual [2016-2017] cumulative incidence: 6.0%). Additionally,
162 this cohort conducted annual visits to the participants' residences to collect blood samples to
163 determine dengue seroconversion (losses to follow-up: 6.5% and 3.2% during the first and second

164 year, respectively), a strategy that allowed for the estimate of attack rates to be calculated for
165 chikungunya (22%) and ZIKV (34%) during the outbreaks of 2015 and 2016, respectively(11).

166

167

168 *Abordando Áreas Endémicas del Dengue para el Estudio de su Severidad (AEDES) Cohorts*

169 The AEDES cohorts' data is an amalgamation of three independent studies, which were assembled
170 with the aim of developing and validating diagnostic and prognostic algorithms for DENV funded
171 by the same national agency above, Minciencias. The first two cohorts were initiated and
172 conducted during endemic periods (2003-2004, n=500; and 2006-2008, N=705) and a third one
173 during an epidemic (2009-2011, N=2,004). These studies shared similar enrolment protocols:
174 febrile patients with clinically suspected dengue were recruited at the point of care. Whereas the
175 first two cohorts were conducted in outpatient settings of Bucaramanga, the third (the AEDES
176 cohort– see Table 2 for clarification) was a multicenter study that enrolled and followed
177 individuals in both outpatient (Bucaramanga, Barranquilla, and Palmira) and inpatient
178 (Bucaramanga, Cali, Neiva, and Palmira) settings. Patients in the first two cohorts came in for
179 follow-up visits one to seven days via clinical and laboratory assessments (median follow-up: 4
180 and 3 days, respectively), with a convalescent blood sample taken approximately 2 weeks after
181 disease onset. The third study which enrolled participants from outpatient and inpatient settings
182 had median follow-up times of 3 and 2 days, respectively. Dengue infection was defined as an
183 ELISA IgM/IgG seroconversion or four-fold increase in titers in paired samples or virus isolation
184 (acute sample) in the first two cohorts, and as an ELISA IgM/IgG seroconversion or four-fold
185 increase in paired samples, or positive ELISA NS1 or PCR, or viral isolation (acute sample) in
186 AEDES(12).

187

188

189 *Efficacy and Safety of a Novel Tetravalent Dengue Vaccine in Healthy Children and Adolescents*
190 *Aged 9 to 16 years in Latin America - placebo arm (CYD15)*

191 Participants in the CYD15 were healthy children and adolescents (9-16 years old) who were
192 recruited from a placebo arm of a randomized control trial funded by Sanofi Pasteur and had the
193 primary goal of evaluating the efficacy of the chimeric yellow fever–dengue–tetravalent (CYF-
194 DTV) dengue vaccine. The CYD15 study was a multicenter, placebo-controlled, randomized,
195 observer-blind phase III DENV vaccine efficacy clinical trial, examining the efficacy of a vaccine
196 to prevent symptomatic virologically-confirmed dengue cases in infants. Participants were
197 randomly assigned, in a 2:1 ratio, to receive three doses of the Recombinant, live, attenuated,
198 tetravalent dengue vaccine, CYD-TDV, (treatment group) or placebo (0.9% sodium chloride;
199 control group) within 1 year. Participants included in this meta-cohort were healthy children and
200 adolescents between 9-16 years old from the placebo arm living in Bucaramanga, Colombia.
201 Volunteers were invited to participate through contacts with schools in the metropolitan area,
202 referred by relatives of participants, or recruited by community leaders. Participants were followed
203 through biweekly phone calls. In case of identifying any febrile episode, participants were asked
204 to provide blood samples to perform virological confirmation and ELISA (NS1, IgM/IgG) testing
205 for dengue infection as well as hemogram and hepatic function tests. Additional ELISA (IgM/IgG),
206 hemogram and hepatic function tests were repeated in a convalescent sample collected up to 21
207 days after fever's onset(13).

208

209 *Pediatric Dengue Cohort Study (PDCS)*

210 The PDCS was established as a community-based cohort in District II of Managua, Nicaragua in
211 2004. The cohort was initially established to study DENV transmission and to characterize

212 symptoms and disease spectrum. It has since evolved to study the virologic and immunologic
213 determinants of response to sequential DENV and ZIKV infections, epidemiological risk factors
214 for infection and disease, and expanded to include other arboviruses, including CHIKV(14,15).
215 The cohort has been funded by a variety of sources, including the US National Institutes of Health
216 (NIH) National Institute of Allergy and Infectious Disease (NIAID), the Pediatric Dengue Vaccine
217 Initiative of the International Vaccine Institute (IVI), the Bill and Melinda Gates Foundation,
218 among others. PDCS is a community-based cohort study which enrolled 2-9-year-olds in August
219 2004. Participants were originally invited to remain in the study until their 12th birthday, but the
220 restriction increased to 15 in 2007, and 18 in 2019(16). Each year, new two-year old children are
221 enrolled, and additional replacement enrollment is performed as needed in the older age groups to
222 maintain the cohort's age structure. At any given time, there are roughly 3,800 - 4,100 children
223 actively participating in the PDCS(17,18). With the introduction of chikungunya virus (CHIKV)
224 and ZIKV into Latin America and specifically Nicaragua, CHIKV and ZIKV were added to the
225 PDCS in 2014 and 2015, respectively. Visits are divided into 4 categories (A-D) based on
226 symptomatology(19). Category A cases include fever plus symptoms and signs of suspected
227 dengue (WHO's case definition). Category B cases are undifferentiated febrile illnesses. Category
228 C cases are fevers with a non-arboviral diagnosis (e.g., influenza, UTI), and category D cases are
229 non-febrile cases. With the introduction of ZIKV, the D case category was divided into two
230 subsets, D cases with ZIKV-like symptoms such as red eyes and rash and D cases without ZIKV-
231 like symptoms(20).
232
233 Category A, B, and ZIKV-like D cases are tested for DENV, ZIKV, and CHIKV using RT-PCR
234 and serological testing. Acute (1-4 days post-onset of symptoms) and convalescent (14-21 days)

235 samples are collected from all suspected cases. An additional sample is collected at day 4-6 from
236 RT-PCR-confirmed DENV and ZIKV cases for immunological studies(21). Each year, a healthy
237 blood sample (serum/plasma, and PBMCs prepared from a subset) is collected and used (a) to
238 detect arbovirus infections that may not have been apparent but may have occurred throughout
239 the year and (b) for additional immunological studies. Data on socioeconomic factors,
240 demographics, and medical history are collected at enrollment and are updated annually. During
241 clinical visits, detailed information on symptoms and symptom onset is collected(22).

242

243 *Pediatric Dengue Hospital-Based Study (PDHS)*

244 This cohort was founded in 1998 to investigate the clinical, immunological, and viral risk factors
245 for severe DENV, assessing biomarkers, and studying immune responses over time. This study
246 has been supported by the NIAID at NIH through various mechanisms. The Pediatric Hospital-
247 Based Study began in 1998 and enrolls children ages 6 months to 14 years who present to the
248 Hospital Infantil Manuel de Jesus Rivera (HIMJR) with suspected dengue (< 7 days from illness
249 onset)(23). Both in-patient and out-patient suspected cases are eligible for enrollment. Upon
250 enrollment, a complete physical exam is performed, and medical history is collected. Participants
251 are followed throughout the acute phase of their illness and data including vital signs, symptoms,
252 and treatment are recorded daily. Blood samples for complete blood counts, molecular, serological,
253 and virological testing are collected daily for the first three days. An additional convalescent
254 sample is collected 14-21 days after enrollment. A longitudinal arm of this study, for those
255 participants who consent, collects samples 3-, 6-, 12-, and 18-months post-illness for
256 immunological studies. The protocol was amended in 2014 to include CHIKV and again in 2016
257 to include ZIKV.

258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280

International Research Consortium on Dengue Risk Assessment, Management and Surveillance (IDAMS)

The primary objective of the IDAMS Cohort was to evaluate warning signs and predictors or biomarkers associated with progression to severe dengue in order to facilitate triage efforts. Funding was provided by the EC’s Seventh Framework Program. The IDAMS study (2011- 2016) was a prospective multi-center acute febrile illness study conducted in Vietnam, Cambodia, Malaysia, Indonesia, Brazil, Venezuela, and El Salvador. Sites recruited participants who with a history of fever for ≤ 72 -84 hours (site-dependent), presenting with clinical symptoms suggestive of dengue (in patients >5 years of age). Patients were excluded if a) they presented with severe dengue at enrolment, b) a clinician judged the patient was unlikely to attend daily outpatient follow-up visits, or c) the clinical presentation strongly suggested a diagnosis other than dengue (e.g., pneumonia, otitis, etc.). Only the data from Latin America were considered for the meta-cohort. However, as the data structure is homogenous, other study locations could be added in the future. The study design was described before in detail (24,25). In brief, patients with a history of fever for less than 72-84 hours (site dependent) and suggestive of DENV were recruited in outpatient clinics across the participating sites and followed up daily for a maximum of 6 days or until afebrile for 48h, with a final follow-up visit around day 10-14 of illness. Daily follow-up included physical examination as well as simple laboratory investigations such as full blood count. Dengue infection was defined as confirmed positive by either PCR or ELISA NS1 result in acute sample.

281 *Cohort of Symptomatic Pregnant Women*

282 In 2012, a prospective cohort for dengue surveillance in mother–infant pairs was established within
283 the Manguinhos, Rio de Janeiro area. In 2015, however, most of these were later identified as
284 ZIKV cases. To identify these ZIKV cases in the Rio de Janeiro population, the pregnancy cohort
285 study was modified to enroll women who presented with a rash at any week of gestation. It was
286 supported by the Department of Science and Technology (Departamento de Ciência e Tecnologia–
287 DECIT) of the Brazilian Ministry of Health (Ministério da Saúde) and funded by the Coordination
288 of the Improvement of Higher Level Personnel (Coordenação de Aperfeiçoamento de Pessoal de
289 Nível Superior-CAPES); the Bill and Melinda Gates Foundation, Grand Challenges Explorations;
290 and the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of
291 Health. Brazil’s Symptomatic Women cohort offered enrollment between September 2015 to
292 September 2016 to pregnant women who attended the acute febrile illness clinic at the Oswaldo
293 Cruz Foundation and who presented with a rash that had developed in the previous 5 days, with or
294 without an associated fever. Laboratory data was collected after enrollment(26). Weekly follow-
295 ups occurred over the phone, and clinical and laboratory follow-up occurred within 30 days of
296 enrollment and were referred for fetal ultrasonography follow-ups at three timepoints: before 20
297 weeks of gestation, between 20-30 weeks of gestation, and after 30 weeks of gestation. ZIKV
298 infection was defined as positive RT-PCR, in acute samples of blood or urine.
299

300 Table 2: Summary of studies in the Meta-Cohort

Country	Study	Years Study was Conducted	Included pediatric cohorts	Study Status	N	Adherence*								
						Mean age (yr)	Annual Loss to follow-up (%)	Followed			Lost			
								Mean age (yr)	Female (%)	Confirmed Dengue (%)	Mean age (yr)	Female (%)	Confirmed Dengue (%)	
COL	Diagnosis Cohort (Evaluation of the diagnostic accuracy and usefulness of rapid tests for early diagnosis of dengue)	2019-2021	x	Completed	28	20.5	10.7	21.1	60	40	8.0	33.3	33.3	
	Prognostic Cohort (Immune Mechanisms of Pathogenesis in Patients with Dengue Infection)	2019-2021	x	Completed	256	13.4	5.5	13.4	47.1	100	13.1	35.7	100	
	PHBDC (Piedecuesta's household-based dynamic cohort: Identification of age groups to be prioritized for vaccination in a population of children and adolescents)	2015-2018	x	Completed	2730	12.6	10.1	14.6	51.2	6.0	12.3	48.5	NA	
	A E D E S c o h o r t s	Identification of prognostic markers of severity in dengue	2003-2004	x	Completed	500	24.7	NA	24.7	51.2	39.0	NA	NA	NA
		Validation of a clinical definition for dengue and evaluation of its usefulness to identify early conditions associated with hospitalization	2006-2008	x	Completed	705	26.5	NA	26.5	42.8	32.3	NA	NA	NA
	A E D E S c o h o r t s	AEDES (Abordando Áreas Endémicas del Dengue para el Estudio de su Severidad). Colombian multicentric study.	Outpatient	x	Completed	1008	21.7	NA	21.7	48.2	38.6	NA	NA	NA
			Inpatient		Completed	996	14.8	NA	14.8	47.3	41.2	NA	NA	NA
	CYD15 (Placebo Arm): Efficacy and safety of a new tetravalent dengue vaccine in healthy children and adolescents aged 9 to 16 years in Latin America	2011-2017	x	Completed	484	12.0	13.0	12.0	51.8	8.3	12.0	48.0	3.2	
NICA	PDCS (Pediatric Dengue Cohort Study)	2004- ongoing	x	Ongoing	9,699	6	4.3 [#]	7.9	49.00	11.5	NA	NA	NA	
	PDHS (Prospective Hospital-Based Study of Dengue Classification, Case Management, and Diagnosis)	1998- ongoing	x	Ongoing	2,659	13	7.03	8.59	48.10	53.7	NA	NA	NA	
IDAMS Latin America	IDAMS (International Research Consortium on Dengue Risk Assessment, Management and Surveillance)****	2012-2016		Completed	1625	19.3	2.7*	19.0	47.6	28.8	13.4	86.4	NA	
BRA	Cohort of Symptomatic Pregnant Women	2015- 2016		Completed	383	6.27	13.6	29.77	86.40	0	27.56	13.60	0	

* Excluded (not lost) because: (i) > 84 hours from disease onset (N=39); (ii) disease onset unknown (N=5); (iii) laboratory diagnosis not available (N=53). 7 patients fulfilled more than one criterion. NA: Not available.

**** Venezuela and Brazil sub-cohorts (Total N enrolled in all sites including Asia = 7428)

[#]Average annual rate in longstanding cohort studies

301
302
303

304 Table 3. Summary of ethics approvals and consent among patients

Study		Did you obtain ethics approval for this study?	Start date	End date	Type of informed consent obtained	Does the study include minors?	Was Informed parental consent obtained for minors?	Was informed consent obtained for adults?
Col: Diagnosis Cohort: Evaluation of the diagnostic accuracy and usefulness of rapid tests for early diagnosis of dengue		Yes	January 23, 2020	June 01, 2020	Written	Yes	Yes	Yes
Col: Prognosis: Immune Mechanisms of Pathogenesis in Patients with Dengue Infection		Yes	March 18, 2019	February 20, 2020	Written	Yes	Yes	Yes
Col: Piedecuesta's household-based dynamic cohort. Identification of age groups to be prioritized for vaccination in a population of children and adolescents.		Yes	June 1, 2015	December 30, 2018	Written	Yes	Yes	Yes
Col:AEDES Cohorts	Identification of prognostic markers of severity in dengue	Yes. A copy of IRB approval obtained in 2003 was requested.	April 01, 2003	March 30, 2005	Written	Yes	Yes	No
	Validation of a clinical definition for dengue and evaluation of its usefulness to identify early conditions associated with hospitalization	Yes. A copy of IRB approval obtained in 2005 was requested.	February 01, 2006	April 30, 2008	Written	Yes	Yes	NA
	"AEDES: Abordando Áreas Endémicas del Dengue para el Estudio de su Severidad". Colombian multicentric study.	Yes	May 01, 2009	May 30, 2011	Written	Yes	Yes	NA
CDY15: Efficacy and safety of a new tetravalent dengue vaccine in healthy children and adolescents aged 9 to 16 years in Latin America		yes	Jun 15, 2011	Feb 28, 2018	Written	Yes	Yes	Yes

Pediatric Dengue Cohort Study (PDCS) (Nicaragua)	Yes	August 31, 2004	ongoing	Informed consent, parental; verbal assent	Yes	Yes	Yes
Pediatric Dengue Hospital-Based Study (PDHS) (Nicaragua)	Yes	August 4, 2005	ongoing	Informed consent; verbal assent; written assent	Yes	Yes	Yes
IDAMS (Carabobo, Venezuela)	Yes	September 27, 2013	November 14, 2016	Written	Yes	Yes	Yes
IDAMS (Rio de Janeiro, Brazil)	Yes	April 6th, 2015	May 9th, 2016	Written	Yes	Yes	Yes
Cohort of Pregnant Women (Brazil)	Yes	September 1 st , 2015	May 30 th , 2016	Written	Yes	Yes	Yes

306

307 Retrospective and Prospective Harmonization 308 Efforts of the Meta-Cohort

309 Individual patient data (IPD) meta-analyses (MA) are considered the gold standard for
310 meta-analyses(27)The strength of conducting IPD-MAs, as opposed to standard meta-analyses
311 using effect estimates, is the opportunity to control for baseline heterogeneity

312 Knowing that we had a collective wealth of arbovirus data, we initially set out to find a
313 common data dictionary for all cohorts within the ReCoDID consortium. This effort included
314 trying to find common variables across all zika cohorts, then across dengue cohorts, as well as any
315 chikungunya patients. As mentioned in the literature, there has not been a gold standard method
316 for data harmonization, but we eventually found the Maelstrom Guide to Rigorous Retrospective
317 Harmonization (28). The first step of the guide is to define a research objective for harmonization,
318 which allowed us to focus on the dengue studies based on a research question that aimed to identify
319 future flavivirus clinical epidemiology in settings where we can infer the past history of infections.
320 Once we had a focused research question, the next step was implementing a well-defined structure
321 of dimensions (endpoints/confounders/exposure) and domains to tackle from a medical
322 perspective. These dimensions guided us in establishing a set of medical meetings where we
323 identified, prioritized and defined in a semantic manner the variables to include in the master data
324 dictionary as indicated in the methodology used and explained in the Fig 1.

325 Fig 1. Harmonization process.

326

327 Then we compared each study data dictionary against the reconciled master data
328 dictionary establishing a harmonization potential based on a predefined structure of conventions

329 that gave us a complete set of variables with different levels of potential use for harmonization,
330 this can be identified in the Fig 2.

331 Fig 2. Harmonization potential

332

333 Retrospective harmonization is generally resource-intensive (29,30), as we can attest to. In
334 order to facilitate future IPD-MAs, we are in the process of creating a Case Report Form which
335 we then standardized according to the CDISC data standard. The hope is that future cohorts
336 conducting arbovirus research will be able to prospectively implement this CRF, decreasing, or
337 removing entirely, the effort and cost of retrospectively harmonizing data to be pooled in an IPD-
338 MA.

339

340 What has it found? Key findings and publications

341 Key findings include: 1) low serum 25(OH)D concentrations in patients predicted the
342 progression of dengue fever(31); 2) the highest risk group for severe dengue being patients with
343 preexisting anti-DENV antibody titers; the same study showing a preventative effect in those
344 patients who have a (very) high level of antibody titers(32); and 3) the development, validation,
345 and evaluation of the usefulness of a scale for the prediction of disease severity among confirmed
346 cases of dengue (32%, 39%, and 41% for the three cohorts).

347

348

349

350

351 What are the main strengths and weaknesses of the 352 combined meta-cohort effort?

353 A technical strength of the Consortium's efforts is the retrospective harmonization of participant-
354 level data which has been completed according to the Maelstrom group's recommendations for
355 best practice(28), which outlines steps 0 (define the research question) to 5 (disseminate and
356 preserve final harmonization products). Maelstrom Research guidelines' Step 0 is why the
357 ReCoDID Consortium's data outlined in this profile is primarily focused on dengue-related
358 outcomes despite collecting data related to viruses. Another strength is the use of CDISC/SDTM
359 standards as reference for the definition of the core variables harmonized (33), which enables
360 interoperability between the meta-cohort and other studies that apply CDISC SDTM. The
361 harmonization (processing data collected by studies und a common variable format), or
362 standardization (use of a data standard (CDISC, e.g.) to define the core variables format to be
363 generated, required to conduct individual-level meta-analyses can be extremely resource-
364 demanding. Going forward, and in order to minimize this burden and improve individual-level
365 patient data meta-analyses (IPD-Mas), the ReCoDID Consortium recommends the creation of a
366 standard case report form (CRF) for acute viral syndrome (arbovirus) research, which includes the
367 features of the overlapping clinical syndromes associated with the most important arboviruses
368 (e.g., DENV, ZIKV, CHIKV, etc.), to be used by all partner studies in the future. Potentially, this
369 CRF will include modules for different severe disease manifestations that can be adapted to the
370 local situation (i.e., bleeding module, neurological module, liver pathology module).

371
372 Another strength of the combined cohort is that it covers different countries and partner sites across
373 Latin America - each of which experienced slightly different histories of DENV serotypes,

374 CHIKV, and ZIKV infections over the last decades. The resulting ‘experiment of nature’
 375 represents a population immune landscape that we now would like to prospectively follow with
 376 future cohorts, considering the potential of immunological interaction between related flaviviruses
 377 (e.g., DENV1-4 and ZIKV). In Fig 3 and Fig 4, created using R (4.0.5)(34), we present DENV,
 378 ZIKV, and CHIKV activity over time in the respective countries and partner sites. This combined
 379 cohort is a first step towards the direction of a multicentric arbovirus cohort, which needs to be
 380 complemented with advanced technology assessing immunological history, additional investments
 381 in future harmonization, and standardization.
 382

Fig 3. DENV serotype distribution over time.	Fig 4. Reported number of cases for DENV, ZIKV and CHIKV over time
a) Pernambuco (Brazil) * b) Ceará (Brazil) Δ c) Rio de Janeiro (Brazil) Δ d) Nicaragua § e) Colombia ∞ f) Venezuela #	a) Pernambuco (Brazil) ∏, Σ, + b) Ceará (Brazil) ∏, + c) Rio de Janeiro (Brazil) ∏, d) Nicaragua § e) Colombia ∞ f) Venezuela ** g) El Salvador
<p>* Data of annual DENV serotypes from PE, provided by: the Central Laboratory of Pernambuco (Laboratório Central de Pernambuco-LACEN PE) Δ Data of annual DENV serotypes from CE and RJ, retrieved from the national online database of the Brazilian Ministry of Health (https://datasus.saude.gov.br/informacoes-de-saude-tabnet/, accessed on 11/8/21 to 11/11/21) ∞ This data was derived from the Colombian National Institute of Health. http://portalsivigila.ins.gov.co/Paginas/Vigilancia-Rutinaria.aspx # Dengue incidence data (1997-2014), Data from: the National Surveillance System of the Venezuelan mandatory notification diseases, the Ministry of Health (http://www.mpps.gob.ve). Data on the proportion of dengue cases per serotype in Aragua, provided by: the Laboratorio Regional de Diagnóstico e Investigación del Dengue y otras Enfermedades Virales (LARDIDEV), Corporación de Salud Aragua, Maracay, Venezuela, and published in Lizarazo Forero, EF (2019). Epidemiology, genetic diversity and clinical manifestations of arboviral diseases in Venezuela. PhD Thesis. University of Groningen, Groningen, the Netherlands. https://doi.org/10.33612/diss.108089934 ∏ Data of annual DENV, ZIKV and CHIKV, reported cases in the states of Ceará (CE), Pernambuco (PE) and Rio de Janeiro (RJ), retrieved from: the national online database of the Brazilian Ministry of Health (https://datasus.saude.gov.br/informacoes-de-saude-tabnet/, accessed on 11/8/21 to 11/11/21), except for CHIKV in 2015-2016 in CE and PE (see below). Data of CHIKV from CE and PE (2015-2016), taken from: Epidemiological Bulletins (EBs) published by the Brazilian Ministry of Health. Σ Data for ZIKV from PE (2015), may be subject to reporting bias. At the time, ZIKV was largely unknown and therefore ZIKV cases might have been classified/notified as dengue cases, de Brito et al. (2016), Data from: Rev Soc Bras Med Trop 49(5):553-558, September-October, 2016 doi:10.1590/0037-8682-0245-2016. + Data of CHIKV from CE and PE (2015-2016), Taken from; Epidemiological Bulletins (EBs) published by the Brazilian Ministry of Health. § PDCS arboviral case data (Oct 2004 - Mar 2021). Image one contains yearly level data of PCR confirmed Dengue cases, and image two contains DENV, CHIKV and ZIKV confirmed symptomatic infections on a monthly basis. ** a) Vincenti-Gonzalez MF, A Tami, EF Lizarazo, ME Grillet (2018), Data from: ENSO-driven climate variability promotes periodic major outbreaks of dengue in Venezuela. Scientific Reports, Apr 10;8(1):5727. doi: 10.1038/s41598-018-24003-z b) Data derived from: Boletines Epidemiológicos de Venezuela, currently found at https://www.ovsalud.org/publicaciones/documentos-oficiales/. c) Data from: Venezuela National EPI-12 notifications, weeks 1-29, (2016). d) PAHO/WHO Epidemiological weekly bulletins at https://www.paho.org/en/documents</p>	

383

384

385

386

387 **Can I access the data? Where can I find out**
388 **more?**

389 Interested parties will be able to access data dictionaries that include information on variables
390 across the datasets via [BioStudies](#)(35). After consultation with each cohort, linked harmonized and
391 curated human cohort data (CE and HDL) is planned to be made accessible through the [European](#)
392 [Genome-phenome Archive \(EGA\)](#)(36) platform to Data Users after their requests are evaluated by
393 the ReCoDID Data Access Committee (DAC). Researchers will be able to openly access the
394 descriptive cohort metadata related to the meta-cohort via the EMBL-EBI [Cohort Browser](#)(37).
395 Cohorts sharing CE and HDL data within the ReCoDID Platform are responsible for obtaining
396 regulatory and ethical approvals at the local, or, in the case of Brazil, national, level for data
397 sharing. Where cohorts' original informed consent forms did not include broad consent for future
398 use of data, ReCoDID worked with cohorts to apply for a waiver of consent. The waivers of
399 consent submitted to the Commission of Ethics in Research, CONEP, Brazil's national ethics
400 regulatory authority. The waivers of consent have been in process for one year and we expect we
401 will need two years to complete the CONEP review.. Samples will not be shared; however, the
402 associated viral sequencing data can be shared and accessed openly at the [European Nucleotide](#)
403 [Archive \(ENA\)](#)(38). Any further associated and shared data types can be shared and retrieved from
404 the relevant [EMBL-EBI resources](#)(39).

405

406 The assessment of political, ethical, administrative, regulatory, and legal (PEARL) issues related
407 to data sharing guided the implementation and activities of ReCoDID to promote ethical data

408 governance and sharing that carefully considers the perspective and context of Low and Middle
409 Income Countries (LMICs). Empirical research, bibliography reviews, and stakeholders'
410 consultations guided ReCoDID members' activities and informed the development of the
411 ReCoDID Data Governance Framework (DGF). The ReCoDID DGF is a high-level normative,
412 organizational, and technical document that describes the goals and principles by which the
413 ReCoDID functions, among them the FAIR (findable, accessible, interoperable, and reusable) data
414 principles. The ReCoDID DGF also implements international standards and best practices for data
415 sharing to promote the public interest and advancement of science. It further describes how the
416 ReCoDID Platform functions and complies with data protection and privacy legislation, mainly
417 the European Union (EU) General Data Protection Regulation (GDPR), and how it relates to other
418 countries' legislations considering international data transfers. The ReCoDID DGF is centred on
419 protecting the rights and interests of Data Subjects (patients and research participants) and different
420 stakeholders that participate in the biomedical innovation ecosystem, including researchers and
421 information technology (IT) professionals that engage in biomedical research and develop the
422 infrastructure and tools that enable health data platforms.

423 ReCoDID acknowledges the challenges to promoting a biomedical innovation system that is
424 transparent, equitable, and participatory, that incorporates LMICs' context-specific perspectives.
425 This is deeply rooted in identifying and overcoming the PEARL barriers related to data sharing
426 mentioned above, as well as the enablers and different strategies to guarantee and implement
427 ethical data sharing and governance. Addressing these PEARL barriers requires creating
428 discussion forums, building research networks, and promoting best practices among academic
429 communities that call for the de-colonization of global health practices and challenge power
430 structures that support and perpetuate them. Therefore, the ReCoDID DGF and ReCoDID

431 Intellectual Property and Open Science Policy incorporate these elements and include issues
432 related to benefit sharing, authorship, attribution, and recognition, as well as the mechanism to
433 implement them.

434 For more information, please contact ena-path-collabs@ebi.ac.uk, with reference to the ReCoDID
435 Project. To share associated human cohort data, please contact the University of Heidelberg
436 (thomas.jaenisch@uni-heidelberg.de).

437

438

439 Funding

440 ReCoDID is supported by the European Union's Horizon 2020 Research and Innovation
441 Programme under Grant Agreement No. 825746 and the Canadian Institutes of Health Research,
442 Institute of Genetics (CIHR-IG) under Grant Agreement N.01886-000. The funders had no role in
443 study design, data collection and analysis, decision to publish, or preparation of the manuscript.

444



445

446

447 Conflict of Interest

448 None declared.

449

450

451

452
453
454

455 Acknowledgements

456
457
458

This work was supported by the ReCoDID Arbovirus harmonization study group:

459 **Janet Achieng** (*Heidelberg Institute of Global Health (HIGH), Heidelberg University*
460 *Hospital, Heidelberg, Germany*)

461 **Sonia Arguello** (*Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brasil*)

462 **Gabriela Maria Marón Alfaro** (*Hospital Nacional de Niños Benjamín Bloom, San Salvador,*
463 *El Salvador*)

464 **Till Bärnighausen** (*Heidelberg Institute of Global Health (HIGH), Heidelberg University*
465 *Hospital, Heidelberg, Germany*)

466 **Bruno Souza Benevides** (*University of the State of Ceara, Fortaleza, Brazil*)

467 **Sarah Bethencourt** (*Departamento de Estudios Clínicos, Facultad de Ciencias de la Salud,*
468 *Universidad de Carabobo, Valencia, Venezuela*)

469 **Andrea Caprara** (*Departamento de Estudios Clínicos, Facultad de Ciencias de la Salud,*
470 *Universidad de Carabobo, Valencia, Venezuela*)

471 **Priscilla M.D.G. César** (*McMaster University, Institute on Ethics & Policy for Innovation*
472 *(IEPI)*)

473 **Guy Cochrane** (*European Molecular Biology Laboratory, European Bioinformatics Institute,*
474 *Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK*)

475 **Monika Consuegra** (*Centro de Atención y Diagnóstico de Enfermedades Infecciosas,*
476 *Bucaramanga, Colombia*)

477 **Fabio Otero** (*Centro de Atención y Diagnóstico de Enfermedades Infecciosas, Bucaramanga,*
478 *Colombia*)

479 **Kerstin Rosenberger** (*Heidelberg Institute of Global Health (HIGH), Heidelberg University*
480 *Hospital, Heidelberg, Germany*)

481 **Anna M. Gajewski** (*Sustainable Sciences Institute, Managua, Nicaragua*)

482 **Nadim Rahman** (*European Molecular Biology Laboratory, European Bioinformatics*
483 *Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK*)

484 **Bladimir Cruz** (*Hospital Nacional de Niños Benjamín Bloom, San Salvador, El Salvador*)

485 **Guillermo Barahona Escobar** (*Hospital Nacional de Niños Benjamín Bloom, San Salvador,*
486 *El Salvador*)

487 **Maria I. Estupiñán** (*Centro de Atención y Diagnóstico de Enfermedades Infecciosas,*
488 *Bucaramanga, Colombia*)

489 **Isabel Fortier** (*Research Institute, McGill University Health Centre, Montreal, Canada*)

490 **Rosa Margarita Gélvez Ramírez** (*Centro de Atención y Diagnóstico de Enfermedades*
491 *Infecciosas, Bucaramanga, Colombia*)

492 **María Fernanda Vincenti Gonzalez** (*University of Groningen, University Medical Center*
493 *Groningen, Department of Medical Microbiology and Infection Prevention, Groningen, The*
494 *Netherlands*)

495 **Peter W. Harrison** (*European Molecular Biology Laboratory, European Bioinformatics*
496 *Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK*)

497 **Suran Jayathilaka** (*European Molecular Biology Laboratory, European Bioinformatics*
498 *Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK*)

499 **Guillermina Kuan** (*Sustainable Sciences Institute, Managua, Nicaragua*)

500 **Manish Kumar** (European Molecular Biology Laboratory, European Bioinformatics Institute,
501 Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK)

502 **Luz Marina Leegstra** (Heidelberg Institute of Global Health (HIGH), Heidelberg University
503 Hospital, Heidelberg, Germany)

504 **Erley Ferlipe Lizarazo** (University of Groningen, University Medical Center Groningen,
505 Department of Medical Microbiology and Infection Prevention, Groningen, The Netherlands)

506 **Luigi Marongiu** (Heidelberg Institute of Global Health (HIGH), Heidelberg University
507 Hospital, Heidelberg, Germany)

508 **Ágnes Molnár** (Heidelberg Institute of Global Health (HIGH), Heidelberg University
509 Hospital, Heidelberg, Germany)

510 **Cesar Narvaez** (Sustainable Sciences Institute, Managua, Nicaragua)

511 **Federico Narvaez** (Sustainable Sciences Institute, Managua, Nicaragua)

512 **Sergio Ojeda** (Sustainable Sciences Institute, Managua, Nicaragua)

513 **Gabriele Rinck** (European Molecular Biology Laboratory, European Bioinformatics Institute,
514 Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK)

515 **Ernesto Pleités Sandoval** (Hospital Nacional de Niños Benjamín Bloom, San Salvador, El
516 Salvador)

517 **Frank Tobian** (Heidelberg Institute of Global Health (HIGH), Heidelberg University
518 Hospital, Heidelberg, Germany)

519 **José Victor Zambrana** (Sustainable Sciences Institute, Managua, Nicaragua)

520 16. **Luana Damasceno** (Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brasil)

521

522

523

524 References

- 525 1. Katzelnick LC, Zambrana JV, Elizondo D, Collado D, Garcia N, Arguello S, et al. Dengue and Zika
526 virus infection in children elicit cross-reactive protective and enhancing antibodies that persist
527 long term. *Sci Transl Med* [Internet]. 2021 Oct 10 [cited 2022 Nov 7];13(614):eabg9478. Available
528 from: [/pmc/articles/PMC8693842/](https://pubmed.ncbi.nlm.nih.gov/3693842/)
- 529 2. Petzold S, Agbaria N, Deckert A, Dambach P, Winkler V, Drexler JF, et al. Congenital abnormalities
530 associated with Zika virus infection—Dengue as potential co-factor? A systematic review. *PLoS*
531 *Negl Trop Dis* [Internet]. 2021 [cited 2022 Nov 7];15(1):e0008984. Available from:
532 <https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0008984>
- 533 3. Katzelnick LC, Montoya M, Gresh L, Balmaseda A, Harris E. Neutralizing antibody titers against
534 dengue virus correlate with protection from symptomatic infection in a longitudinal cohort. *Proc*
535 *Natl Acad Sci U S A* [Internet]. 2016 Jan 19 [cited 2022 Nov 8];113(3):728–33. Available from:
536 <https://pubmed.ncbi.nlm.nih.gov/26729879/>
- 537 4. Katzelnick LC, Coloma J, Harris E. Dengue: Knowledge gaps, unmet needs and research priorities.
538 *Lancet Infect Dis* [Internet]. 2017 Mar 1 [cited 2022 Nov 7];17(3):e88. Available from:
539 [/pmc/articles/PMC5967882/](https://pubmed.ncbi.nlm.nih.gov/26729879/)
- 540 5. Katzelnick LC, Harris E, Baric R, Collier BA, Coloma J, Crowe JE, et al. Immune correlates of
541 protection for dengue: State of the art and research agenda. *Vaccine* [Internet]. 2017 Aug 8
542 [cited 2022 Nov 7];35(36):4659. Available from: [/pmc/articles/PMC5924688/](https://pubmed.ncbi.nlm.nih.gov/26729879/)
- 543 6. ReCoDID | Reconciliation of Cohort data in Infectious Diseases [Internet]. [cited 2022 Nov 7].
544 Available from: <https://recodid.eu/>
- 545 7. Díaz FA, Martínez RA, Villar LA. Criterios clínicos para diagnosticar el dengue en los primeros días
546 de enfermedad. *Biomédica* [Internet]. 2006 Mar 1 [cited 2022 Nov 7];26(1):22–30. Available
547 from: <https://revistabiomedica.org/index.php/biomedica/article/view/1391>
- 548 8. Villar LA, Gélvez RM, Rodríguez JA, Salgado D, Parra B, Osorio L, et al. Biomarcadores pronósticos
549 de gravedad del dengue. *Biomédica* [Internet]. 2013 Aug 1 [cited 2022 Nov 7];33(SUPPL.1):108–
550 16. Available from: <https://revistabiomedica.org/index.php/biomedica/article/view/733/2275>
- 551 9. Villamor E, Villar LA, Lozano A, Herrera VM, Herrán OF. Vitamin D serostatus and dengue fever
552 progression to dengue hemorrhagic fever/dengue shock syndrome. *Epidemiol Infect* [Internet].
553 2017 Oct 1 [cited 2022 Nov 7];145(14):2961. Available from: [/pmc/articles/PMC9152741/](https://pubmed.ncbi.nlm.nih.gov/3693842/)
- 554 10. Villamor E, Villar LA, Lozano-Parra A, Herrera VM, Herrán OF. Serum fatty acids and progression
555 from dengue fever to dengue hemorrhagic fever / dengue shock syndrome. *Br J Nutr* [Internet].
556 2018 Oct 14 [cited 2022 Nov 7];120(7):787. Available from: [/pmc/articles/PMC6150811/](https://pubmed.ncbi.nlm.nih.gov/3693842/)
- 557 11. Cárdenas MIE, Herrera VM, Montoya MCM, Parra AL, Moncayo ZMZ, García JPF, et al.
558 Heterogeneity of dengue transmission in an endemic area of Colombia. *PLoS Negl Trop Dis*
559 [Internet]. 2020 Sep 1 [cited 2022 Nov 7];14(9):e0008122. Available from:
560 <https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0008122>
- 561 12. Rojas EM, Herrera VM, Miranda MC, Rojas DP, Gomez AM, Pallares C, et al. Clinical Indicators of
562 Fatal Dengue in Two Endemic Areas of Colombia: A Hospital-Based Case–Control Study. *Am J*
563 *Trop Med Hyg* [Internet]. 2019 [cited 2022 Nov 7];100(2):411. Available from:

- 564 /pmc/articles/PMC6367622/
- 565 13. L'Azou M, Moureau A, Sarti E, Nealon J, Zambrano B, Wartel TA, et al. Symptomatic Dengue in
566 Children in 10 Asian and Latin American Countries. *N Engl J Med* [Internet]. 2016 Mar 24 [cited
567 2022 Nov 7];374(12):1155–66. Available from: <https://pubmed.ncbi.nlm.nih.gov/27007959/>
- 568 14. Gordon A, Gresh L, Ojeda S, Katzelnick LC, Sanchez N, Mercado JC, et al. Prior dengue virus
569 infection and risk of Zika: A pediatric cohort in Nicaragua. *PLoS Med* [Internet]. 2019 [cited 2022
570 Nov 8];16(1). Available from: /pmc/articles/PMC6342296/
- 571 15. Zambrana JV, Carrillo FB, Burger-Calderon R, Collado D, Sanchez N, Ojeda S, et al.
572 Seroprevalence, risk factor, and spatial analyses of Zika virus infection after the 2016 epidemic in
573 Managua, Nicaragua. *Proc Natl Acad Sci U S A* [Internet]. 2018 Sep 11 [cited 2022 Nov
574 8];115(37):9294–9. Available from: /pmc/articles/PMC6140532/
- 575 16. Standish K, Kuan G, Avilés W, Balmaseda A, Harris E. High Dengue Case Capture Rate in Four
576 Years of a Cohort Study in Nicaragua Compared to National Surveillance Data. *PLoS Negl Trop Dis*
577 [Internet]. 2010 Mar [cited 2022 Nov 8];4(3):e633. Available from:
578 <https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0000633>
- 579 17. Balmaseda A, Standish K, Mercado JC, Matute JC, Tellez Y, Saborío S, et al. Trends in patterns of
580 dengue transmission over 4 years in a pediatric cohort study in Nicaragua. *J Infect Dis* [Internet].
581 2010 Jan [cited 2022 Nov 8];201(1):5–14. Available from:
582 <https://pubmed.ncbi.nlm.nih.gov/19929380/>
- 583 18. Burger-Calderon R, Bustos Carrillo F, Gresh L, Ojeda S, Sanchez N, Plazaola M, et al. Age-
584 dependent manifestations and case definitions of pediatric Zika: a prospective cohort study.
585 *Lancet Infect Dis* [Internet]. 2020 Mar 1 [cited 2022 Nov 8];20(3):371. Available from:
586 /pmc/articles/PMC7085943/
- 587 19. Montoya M, Gresh L, Mercado JC, Williams KL, Vargas MJ, Gutierrez G, et al. Symptomatic Versus
588 Inapparent Outcome in Repeat Dengue Virus Infections Is Influenced by the Time Interval
589 between Infections and Study Year. *PLoS Negl Trop Dis* [Internet]. 2013 [cited 2022 Nov
590 8];7(8):2357. Available from: /pmc/articles/PMC3738476/
- 591 20. Katzelnick LC, Narvaez C, Arguello S, Mercado BL, Collado D, Ampie O, et al. Zika virus infection
592 enhances future risk of severe dengue disease. *Science* [Internet]. 2020 Aug 8 [cited 2022 Nov
593 8];369(6507):1123. Available from: /pmc/articles/PMC8274975/
- 594 21. Waggoner JJ, Balmaseda A, Gresh L, Sahoo MK, Montoya M, Wang C, et al. Homotypic Dengue
595 Virus Reinfections in Nicaraguan Children. *J Infect Dis* [Internet]. 2016 Oct 1 [cited 2022 Nov
596 8];214(7):986–93. Available from: <https://academic.oup.com/jid/article/214/7/986/2388004>
- 597 22. OhAinle M, Balmaseda A, Macalalad AR, Tellez Y, Zody MC, Saborío S, et al. Dynamics of Dengue
598 Disease Severity Determined by the Interplay Between Viral Genetics and Serotype-Specific
599 Immunity. *Sci Transl Med* [Internet]. 2011 Dec 12 [cited 2022 Nov 8];3(114):114ra128. Available
600 from: /pmc/articles/PMC4517192/
- 601 23. Gutiérrez G, Gresh L, Pérez MÁ, Elizondo D, Avilés W, Kuan G, et al. Evaluation of the Diagnostic
602 Utility of the Traditional and Revised WHO Dengue Case Definitions. *PLoS Negl Trop Dis*
603 [Internet]. 2013 [cited 2022 Nov 8];7(8):2385. Available from: /pmc/articles/PMC3749970/
- 604 24. Jaenisch T, Tam DTH, Kieu NTT, Ngoc T, Nam NT, van Kinh N, et al. Clinical evaluation of dengue

- 605 and identification of risk factors for severe disease: protocol for a multicentre study in 8
606 countries. *BMC Infect Dis* [Internet]. 2016 Mar 11 [cited 2022 Nov 7];16(1). Available from:
607 <https://pubmed.ncbi.nlm.nih.gov/26968374/>
- 608 25. Jaenisch T, IDAMS, Sakuntabhai A, DENFREE, Wilder-Smith A, DengueTools. Dengue Research
609 Funded by the European Commission-Scientific Strategies of Three European Dengue Research
610 Consortia. *PLoS Negl Trop Dis* [Internet]. 2013 [cited 2022 Nov 8];7(12). Available from:
611 </pmc/articles/PMC3861113/>
- 612 26. Brasil P, Vasconcelos Z, Kerin T, Gabaglia CR, Ribeiro IP, Bonaldo MC, et al. Zika virus vertical
613 transmission in children with confirmed antenatal exposure. *Nat Commun* [Internet]. 2020 Dec 1
614 [cited 2022 Nov 8];11(1). Available from: </pmc/articles/PMC7360785/>
- 615 27. Thomas D, Radji S, Benedetti A. Systematic review of methods for individual patient data meta-
616 analysis with binary outcomes. *BMC Med Res Methodol* [Internet]. 2014 Jun 19 [cited 2022 Nov
617 24];14(1):1–9. Available from:
618 <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-14-79>
- 619 28. Fortier I, Raina P, van den Heuvel ER, Griffith LE, Craig C, Saliba M, et al. Maelstrom Research
620 guidelines for rigorous retrospective data harmonization. *Int J Epidemiol* [Internet]. 2017 Feb 1
621 [cited 2022 Nov 8];46(1):103. Available from: </pmc/articles/PMC5407152/>
- 622 29. Milhim BHGA, da Rocha LC, Terzian ACB, Mazaro CCP, Augusto MT, Luchs A, et al. Arboviral
623 Infections in Neurological Disorders in Hospitalized Patients in São José do Rio Preto, São Paulo,
624 Brazil. *Viruses* [Internet]. 2022 Jul 1 [cited 2022 Dec 1];14(7). Available from:
625 <https://pubmed.ncbi.nlm.nih.gov/35891468/>
- 626 30. Hortion J, Mutuku FM, Eyherabide AL, Vu DM, Boothroyd DB, Grossi-Soyster EN, et al. Acute
627 Flavivirus and Alphavirus Infections among Children in Two Different Areas of Kenya, 2015. *Am J*
628 *Trop Med Hyg* [Internet]. 2019 [cited 2022 Dec 1];100(1):170–3. Available from:
629 <https://pubmed.ncbi.nlm.nih.gov/30457092/>
- 630 31. Martínez RA, Díaz FA, Villar LA. Evaluación de la definición clínica de dengue sugerida por la
631 Organización Mundial de la Salud. *Biomédica* [Internet]. 2005 Sep 1 [cited 2022 Nov
632 8];25(3):412–6. Available from:
633 <https://revistabiomedica.org/index.php/biomedica/article/view/1365>
- 634 32. Katzelnick LC, Gresh L, Halloran ME, Mercado JC, Kuan G, Gordon A, et al. Antibody-dependent
635 enhancement of severe dengue disease in humans. *Science* [Internet]. 2017 Nov 17 [cited 2022
636 Nov 8];358(6365):929. Available from: </pmc/articles/PMC5858873/>
- 637 33. Standards | CDISC [Internet]. [cited 2022 Nov 8]. Available from:
638 <https://www.cdisc.org/standards>
- 639 34. R Core Team (2020) R A Language and Environment for Statistical Computing. R Foundation for
640 Statistical Computing, Vienna, Austria. - References - Scientific Research Publishing [Internet].
641 [cited 2022 Nov 17]. Available from:
642 <https://scirp.org/reference/referencespapers.aspx?referenceid=3064798>
- 643 35. BioStudies < EMBL-EBI [Internet]. [cited 2022 Dec 1]. Available from:
644 <https://www.ebi.ac.uk/biostudies/>
- 645 36. EGA European Genome-Phenome Archive [Internet]. [cited 2022 Dec 1]. Available from:

646 <https://ega-archive.org/>

647 37. Pathogens [Internet]. [cited 2022 Dec 1]. Available from:
648 <https://www.ebi.ac.uk/ena/pathogens/v2/cohorts>

649 38. ENA Browser [Internet]. [cited 2022 Dec 1]. Available from:
650 <https://www.ebi.ac.uk/ena/browser/home>

651 39. An introduction to EMBL-EBI resources | EMBL-EBI Training [Internet]. [cited 2022 Dec 1].
652 Available from: <https://www.ebi.ac.uk/training/events/introduction-embl-ebi-resources/>

653

654

655

656

Master Definition						Studies	Variables	Harmonization		
domain	Temp Variable names	definition	name	unit	type	Studies	Variables	Harmo Potential Status per Variable (to be completed following the instructions below)	Harmo Potential Status detail per Variable ((to be completed following the instructions below)	Comments on harmonization rule
endpoints	age	* age of the patient at the day of measurement * reported by patient or family or official documents	age	years	Integer	Nica- Cohorte Arbovirus	exp	1	1.2	Transformation: Substact(fecha_consulta - exp)
endpoints	age	* age of the patient at the day of measurement * reported by patient or family or official documents	age	years	Integer	Nica- Estudio Hospitalario	Edad	1	1.2	Transformation: DD-MM-YYYY -> YYYY-MM-DD
endpoints	age	* age of the patient at the day of measurement * reported by patient or family or official documents	age	years	Integer	IDAMS	dm_age	1	1.1	idéntico
endpoints	age	* age of the patient at the day of measurement * reported by patient or family or official documents	age	years	Integer	Col-AedesCohort	Edad	1	1.1	idéntico
endpoints	age	* age of the patient at the day of measurement * reported by patient or family or official documents	age	years	Integer	Col: Diagnosis Cohort	Edad	1	1.1	idéntico
endpoints	age	* age of the patient at the day of measurement * reported by patient or family or official documents	age	years	Integer	Col: Prognosis Cohort	Edad	1	1.1	idéntico
endpoints	age	* age of the patient at the day of measurement * reported by patient or family or official documents	age	years	Integer	Col: Seroprevalence Cohort	x	2	2.1	indisponible
endpoints	age	* age of the patient at the day of measurement * reported by patient or family or official documents	age	years	Integer	Brazil _ cohort of symptomatic pregnant women	x	2	2.1	indisponible

Figure 2. Harmonization Potential

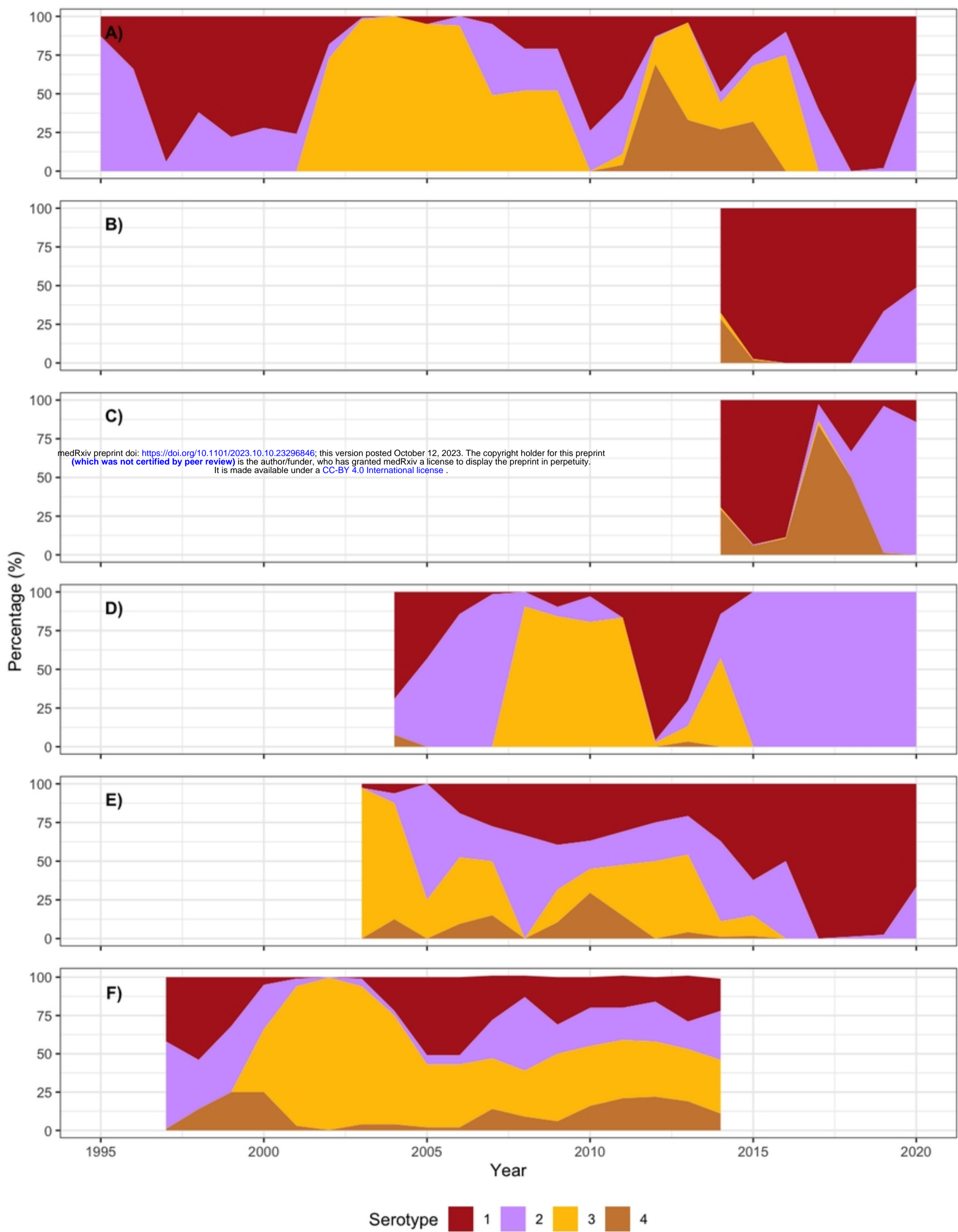
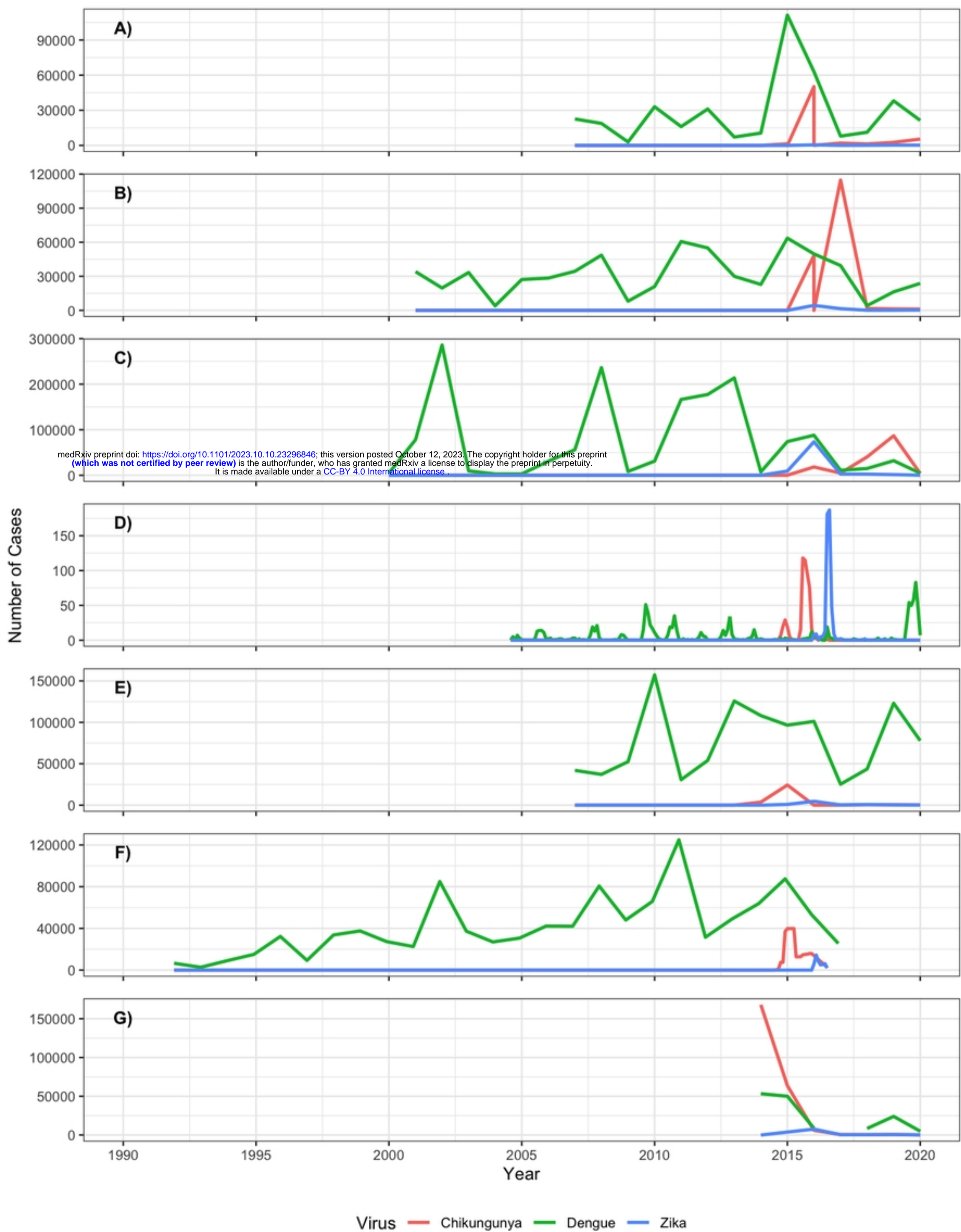


Figure 3. DENV Serotype Distribution



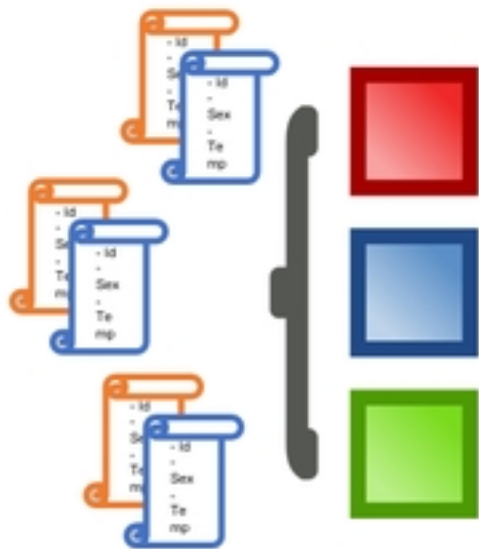
Research Objective

Reconciliation
by Dimension

Data Schema

Harmonization
Rules

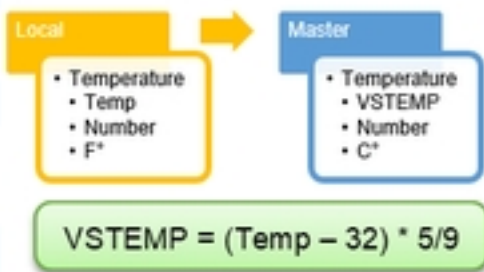
Local datasets
transformation



- Raw variables
- Master definition

Age
• DM-AGE
• Number
Sex
• DM-SEX
• Text
• [F, M]
Temperature
• VS-TEMP
• Number

Each local dataset vs
Data schema



Transformation
process to comply with
master dataset



Figure 1. Harmonization Process