Applied immuno-epidemiological research:

An approach for integrating existing knowledge into the statistical analysis of multiple immune markers

Bernd Genser, MSc, PhD [1 2]

[1] BGStats Consulting, Vienna, Austria
[2] Mannheim Institute of Public Health, Social and Preventive Medicine, University of Heidelberg, Germany
bernd.genser@bgstats.com
www.bgstats.com

PSI One Day Meeting: Immunology Diseases
17th June – Roche Products Limited, 6 Falcon Way, Shire Park, Welwyn Garden City, A171TW

Outline of the presentation

Background:

- Challenges for statistical analysis of immuno-epidemiological data
- The need for an analytical framework of simultaneous analysis of multiple correlated markers

Methods:

- A step-wise integrated data analytic approach
- The application: cytokine data from a large immuno-epidemiological study investigating risk factors of atopic disease and asthma

Results:

- Aggregation of multiple cytokines to immunological summary scores
- Interdependence analysis between summary scores
- Contrast the integrated approach to a traditional regression approach
  - Example: Association testing cytokine patterns vs. specific IgE:

Discussion/Conclusion:

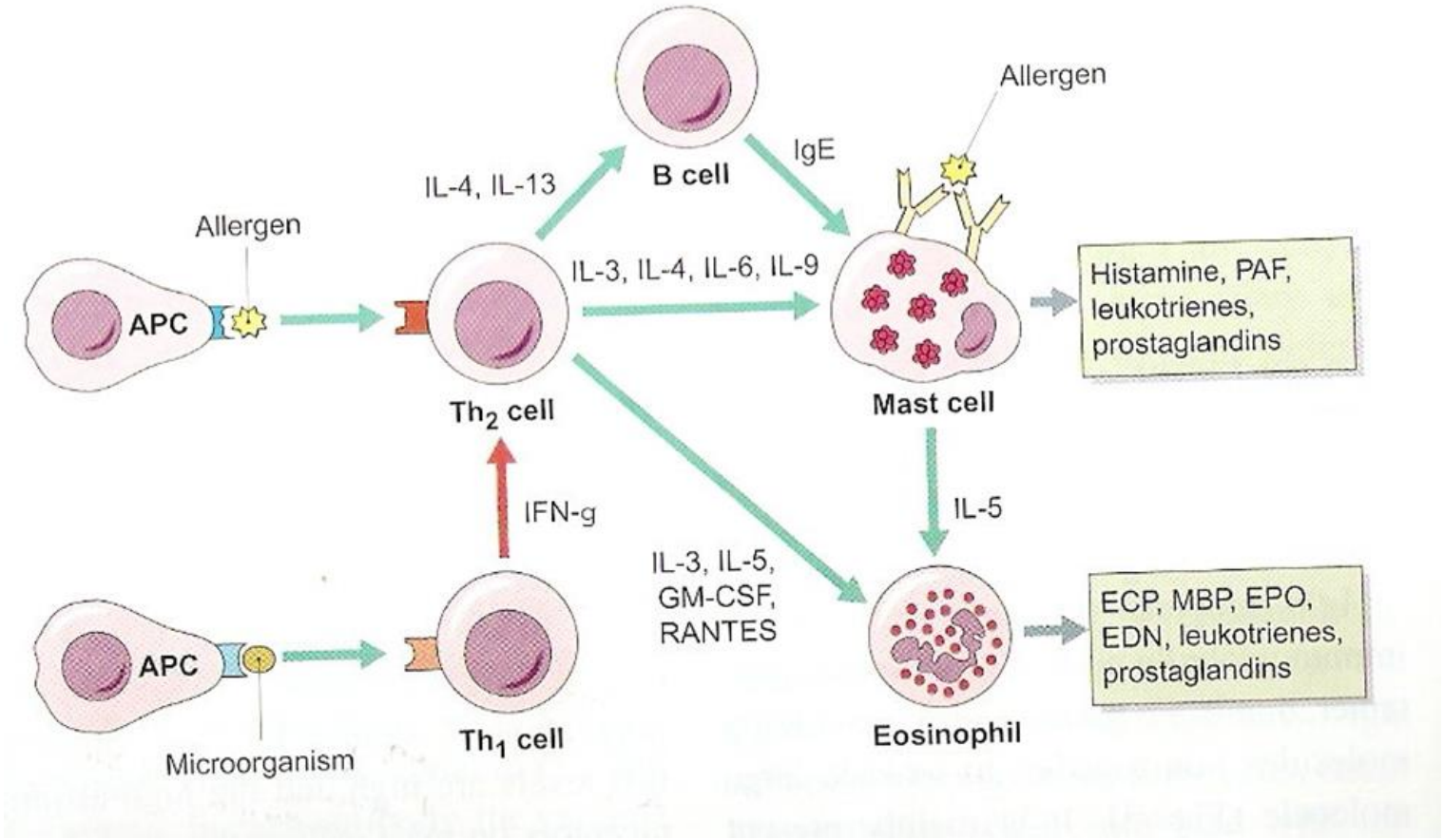
- Interpretation of the findings of integrated data analysis from the application example
- Potential applications of the approach in modern immunological research

BACKGROUND

Motivation for an integrated data analytic approach in modern immunological research

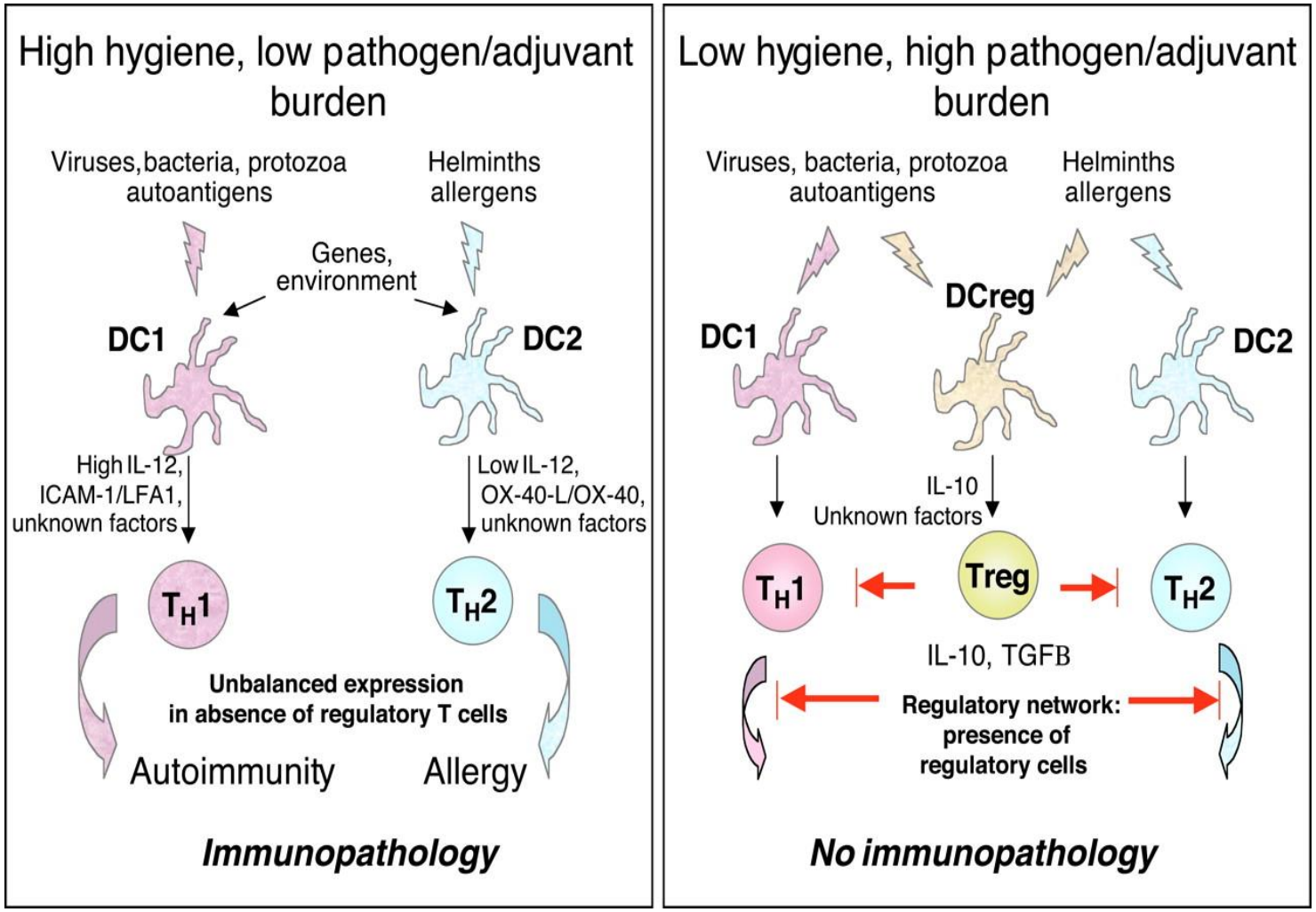
- Advances in modern immunological research: Increasing knowledge about complex mechanism of immunologically mediated diseases, eg. atopy or asthma
- Researchers collect many different immune markers aimed to quantify the presumed underlying immunological mechanisms (eg. Th2 related response)
- Markers are the product of common underlying immunological mechanisms (e.g. produced by the same type of immune cell) and thus often highly correlated
- Advances in immuno-epidemiology:
  - Increasing epidemiological evidence of environmental, social, biological or genetic determinants that affect immunity and thus the risk for immunologically mediated diseases
  - Complex research questions arise investigating relationships between risk factors, immunity and disease
  - Studies result in large and complex datasets including many variables of different types,
  - We expect multiple inter-relationships between the study variables, the collected immune markers are involved in complex causal chains

Example: Immunological concept- Mechanism of allergic inflammation



bg stats

Example: Immunological concepts used in modern immuno-epidemiology

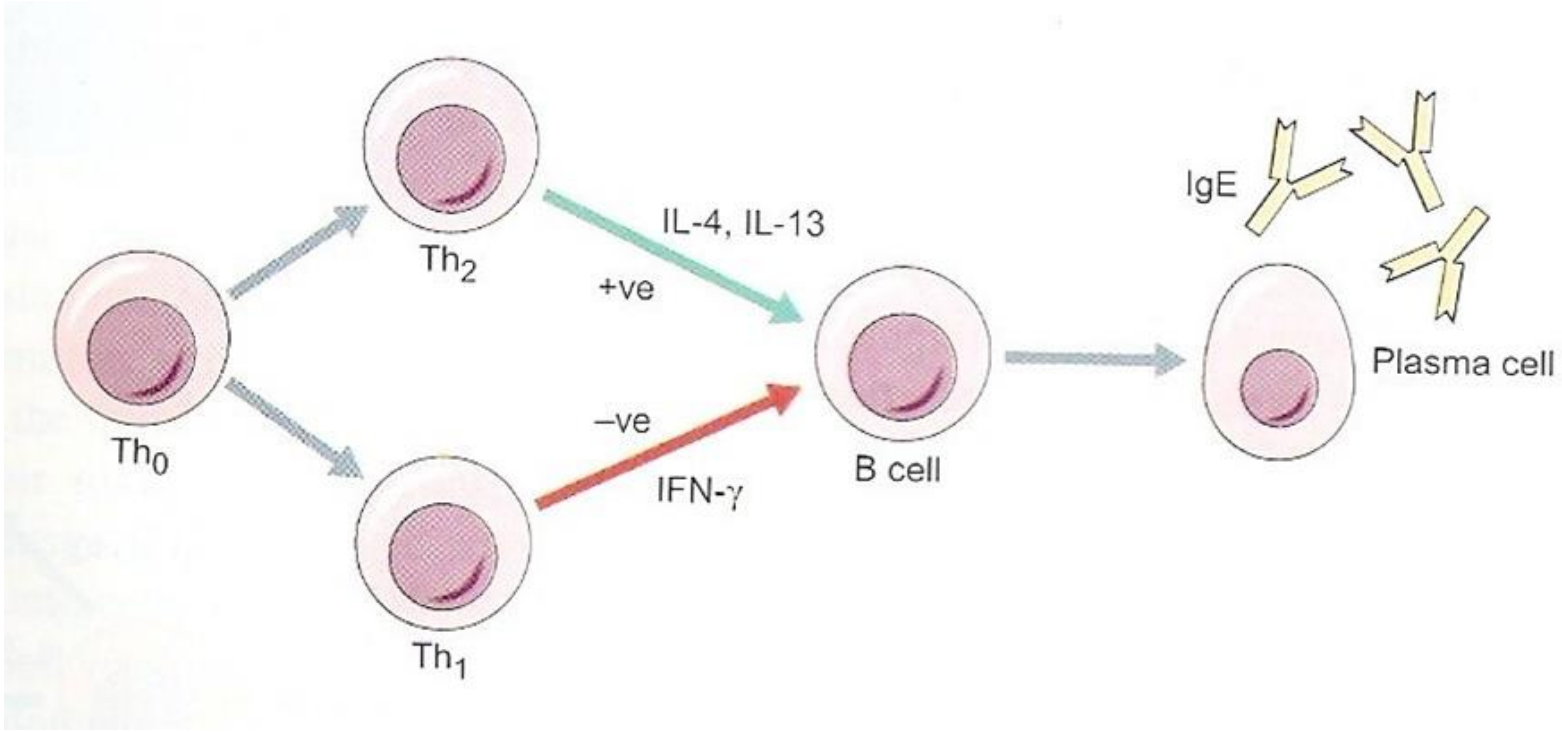


bg stats

Example: Quantifying immunological concepts by immune markers - the Th1/Th2 paradigm and allergic disease

- For example, in allergy research, immunologists collect measurements of different cytokines presumed to result from different immunological concepts involving different immune cells (eg Th1-, Th2- or T-regulatory cells)
- Further, to quantify the individuals‘ potential for different types of immune responses (eg specific vs. non- specific responses) researchers obtain measurements from cell cultures simulating different immunological conditions (eg spontaneous vs. antigen stimulated responses).
- Conceptually, the immunologists expect that the cytokine measures quantify at least in part the immunological concepts involved in the development of atopic diseases, i.e. the Th1/Th2 paradigms, the immune regulatory network and the regulation of specific IgE.

Example: Immunological concept: Th1/Th2 balance and the regulation of specific IgE



bg stats

Peculiarities of immunological data relevant for statistical analysis

- Immunological data show a variety of peculiarities to be addressed in statistical analysis like non-normality of distributions (eg skewness) or the existence of non-detectable values ("non-detects" are concentrations of a marker below the detection limit of an assay).
- Advances in statistical methodological have substantially improved the incorporation of immunological parameters in "classical statistical analysis", such as multivariate regression models:
  - Robust methods to deal with the non-normality of markers
  - Imputation techniques or Kaplan- Meier techniques to deal with "non-detects"
  - Advanced regression models for non-normal data (Tobit- or Quantile regression)
  - Methods for the analysis of repeated immunological measures

Our previous methodological work on statistical analysis of immunological data …

Commentary

Open Access

A guide to modern statistical analysis of immunological data

Bernd Genser*[1,2], Philip J Cooper[3,4], Maria Yazdanbakhsh[5], Mauricio L Barreto[2] and Laura C Rodrigues[2]

Address: [1]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK, [2]Instituto de Saúde Coletiva, Federal University of Bahia, Salvador, Brazil, [3]Centre for Infection, St George's University of London, London, UK, [4]Instituto de Microbiologia, Universidad San Francisco de Quito, Quito, Ecuador and [5]Department of Parasitology, Leiden University Medical Center, Leiden, The Netherlands

Email: Bernd Genser* - bernd.genser@bgstats.com; Philip J Cooper - pcooper@ecnet.ec; Maria Yazdanbakhsh - M.Yazdanbakhsh@lumc.nl; Mauricio L Barreto - mauricio@ufba.br; Laura C Rodrigues - Laura.Rodrigues@lshtm.ac.uk

* Corresponding author

Published: 26 October 2007

BMC Immunology 2007, 8:27 doi:10.1186/1471-2172-8-27

Received: 22 February 2007
Accepted: 26 October 2007

This article is available from: http://www.biomedcentral.com/1471-2172/8/27

bg stats

… resulted in a guide for selecting appropriate techniques for immune markers

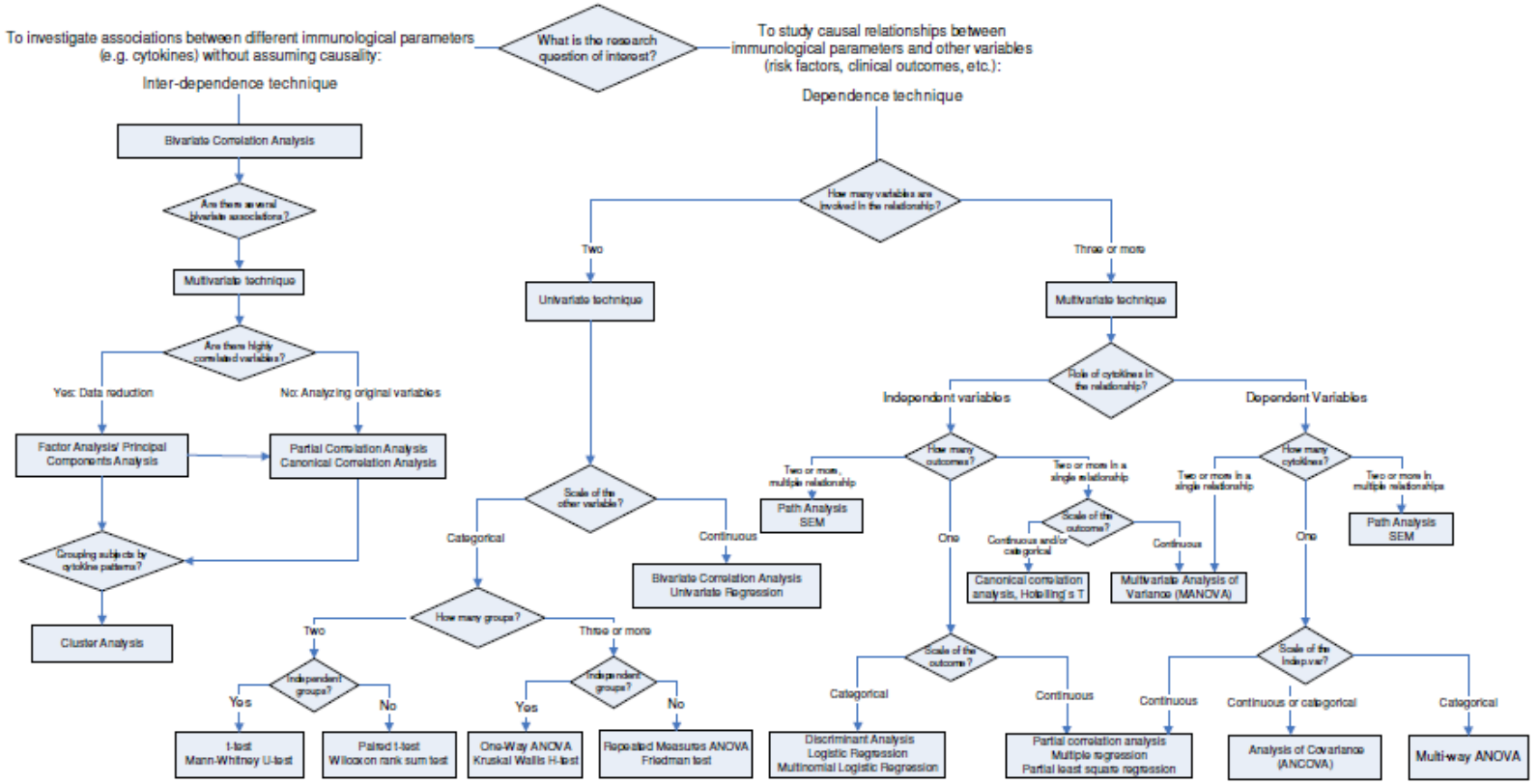


Figure 1
Selecting the appropriate statistical technique for analysis of immunological data.

bg stats

Conceptual frameworks for epidemiological analysis

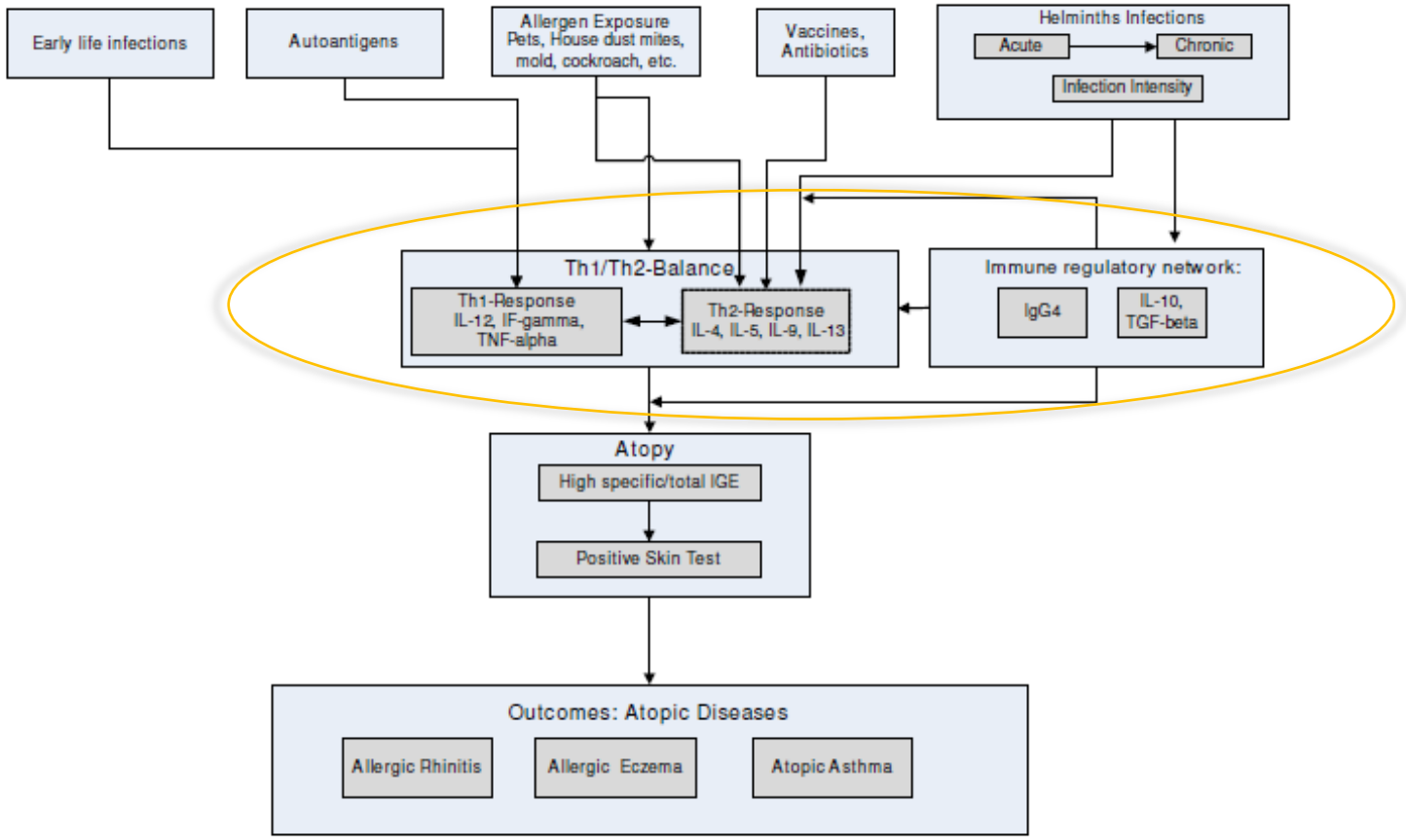
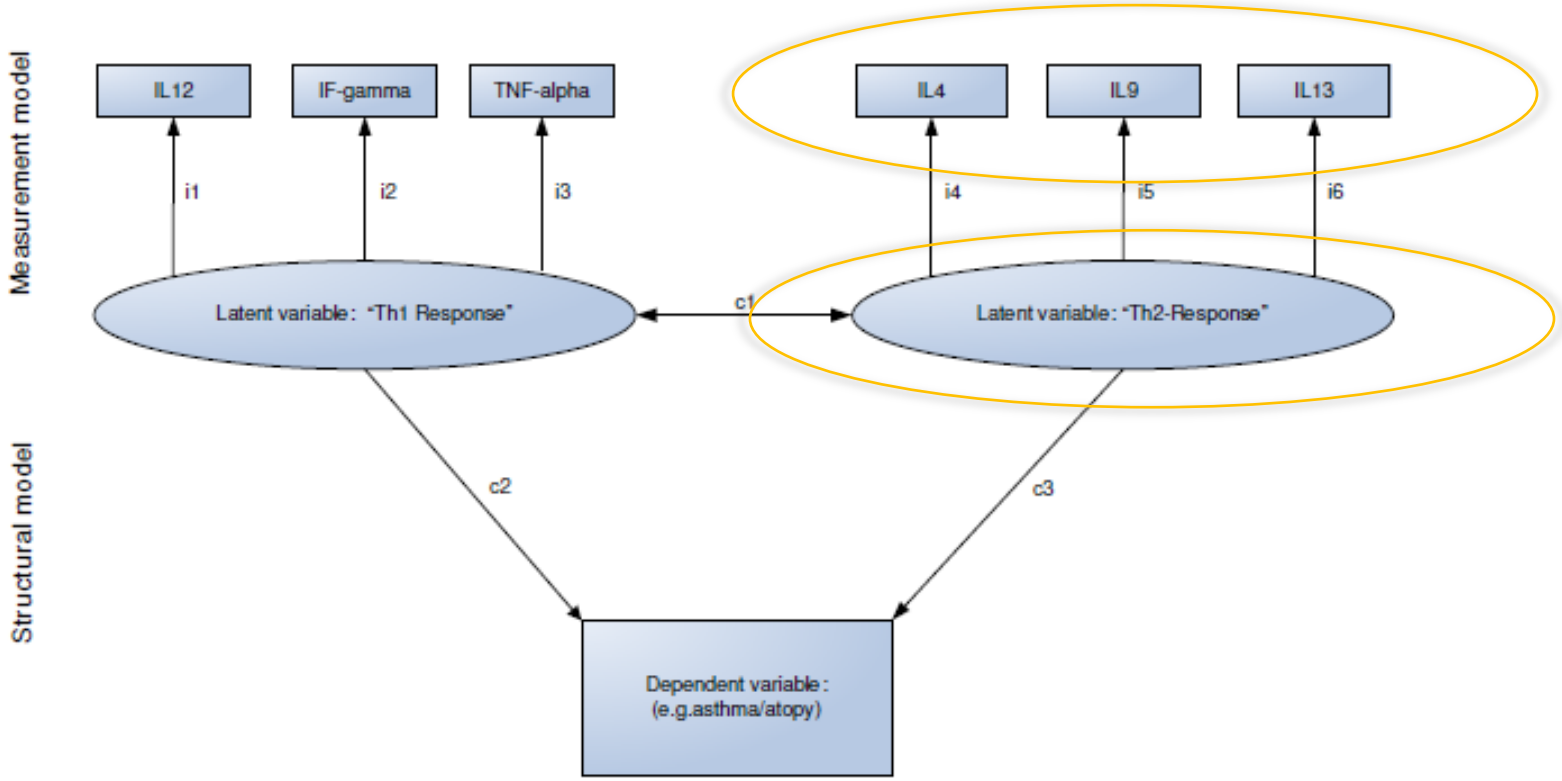


Figure 2
Conceptual framework that specifies multiple associations between potential risk factors, immunological parameters and outcomes (atopy and asthma).

Latent immunological concepts to be quantified by immune markers

bg stats

Structural Equation Models: An approach for modern data analysis of immune markers?



Immune markers

Latent concept

Figure 3
Example path diagram that specifies a structural equation model with two latent variables.

bg stats

Challenges for statistical analysis of multiple correlated markers

There is a strong need for an **analytical approach for simultaneous analysis of multiple immune markers** that are **often correlated** affected by larger immunological mechanisms.

The approach should allow:

- to consider **hierarchical inter-relationships** among the markers e.g. from a multi-level sampling strategy (multiple similar markers are obtained from different immunological experiments)
- to integrate the **researchers' experts knowledge** about the underlying immunological mechanisms

Challenges for statistical analysis of multiple correlated markers (2)

- To address **inter-dependencies among multiple measurements** of the same immune marker,
- To analyze **association patterns** among different markers
- To **aggregate the information** captured in multiple markers to immunological summary scores
- To investigate inter-relationships among the summary scores
- To use the summary scores in epidemiological association analyses with outcomes and/or risk factors (predictors).

METHODS

Methods - Overview

- Part 1: We present an **analytical framework approach** for the **statistical analysis of multiple immune markers** clustered at three functional levels.
- Part 2: We illustrate the **application of the approach to cytokine data from a large immuno-epidemiological study** (SCAALA Salvador) conducted to investigate risk factors and immunological pathways for atopic diseases and asthma.

The application: the SCAALA study

SCAALA: A research programme aimed to investigate the impact of **S**ocial **C**hanges on **A**llergy and **A**sthma in **L**atin **A**merica

Dataset: Multiple immunological markers (cytokines, IgE) from children enrolled in a large immuno-epidemiological study (SCAALA Salvador)

General research question of SCAALA:

- To identify risk factors for atopic diseases and asthma

Specific research questions of SCAALA immunological analysis:

- To **quantify the major presumed immunological mechanisms of atopy** (Th1/Th2 balance, immune regulatory network, specific IgE)
- To **identify factors on different causal levels** (environment, social, genetic etc) that affect the immune profiles and to quantify how changes in immunity affect the risk of atopic disease and asthma

The Salvador-SCAALA study protocol

Study protocol

Open Access

Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in Northeastern Brazil (Salvador-SCAALA study)

Mauricio L Barreto*[1], Sergio S Cunha[1], Neuza Alcântara-Neves[2], Lain P Carvalho[3], Álvaro A Cruz[4], Renato T Stein[5], Bernd Genser[1], Philip J Cooper[6] and Laura C Rodrigues[7]

Address: [1]Instituto de Saúde Coletiva, Universiade Federal de Bahia, Salvador, Brazil, [2]Instituto de Ciências da Saúde, Universidade Federal de Bahia, Salvador, Brazil, [3]Centro de Pesquisas Gonçalo Moniz – FIOCRUZ, Salvador, Brazil, [4]Centro de Enfermidades Respiratórias, Faculdade de Medicina, Universidade Federal de Bahia, Salvador, Brazil, [5]Department of Pediatrics, School of Medicine, Pontifica Universidade Católica, Porto Alegre, Brazil, [6]Instituto de Microbiologia, Universidad San Francisco de Quito, Quito, Ecuador and [7]Department of Epidemiology and Populations health, London School of Hygiene and Tropical Medicine, London, UK

Email: Mauricio L Barreto* - mauricio@ufba.br; Sergio S Cunha - cunhatmt@hotmail.com; Neuza Alcântara-Neves - neuza@ufba.br; Lain P Carvalho - lain@cpqgm.fiocruz.br; Álvaro A Cruz - cruz.aa@terra.com.br; Renato T Stein - rstein@pucrs.br; Bernd Genser - bernd.genser@bgstats.com; Philip J Cooper - pcooper@ecnet.ec; Laura C Rodrigues - Laura.rodrigues@lshtm.ac.uk

* Corresponding author

Published: 23 June 2006

BMC Pulmonary Medicine 2006, **6**:15 doi:10.1186/1471-2466-6-15

Received: 04 May 2006
Accepted: 23 June 2006

This article is available from: http://www.biomedcentral.com/1471-2466/6/15

bg stats

Challenges of statistical analysis in SCAALA

Specific objectives:

- To systematically aggregate the information captured in multiple correlated cytokine measurements to immunological summary scores
- The scores should reflect the major immunological mechanisms presumed to underlie atopic disease (Th1/Th2 balance and immune regulatory network).
- The scores should enhance power and validity of any epidemiological association analysis better quantifying immunity than the original markers.

In detail we sought to investigate:

- how epidemiological factors affect immune patterns
- how changes in immune patterns affect the risk of atopic disease (IgE, SPT+) and asthma (wheezing), i.e. "immunological mediation" of risk factor effects

The study population

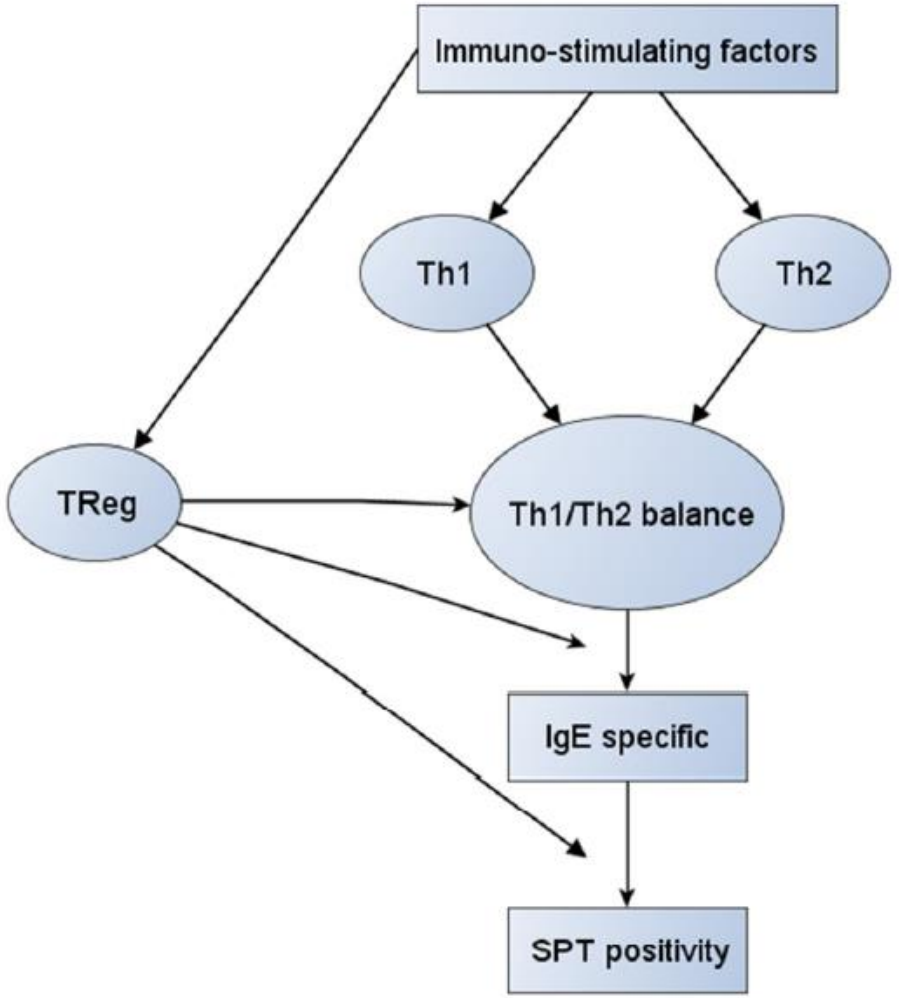
- **Population of SCAALA Salvador:** N=1445 children **aged 4-11 years** living in the city of Salvador, a large urban developing center in Northeast Brazil with a population of 2.8 million and a high prevalence of asthma symptoms such as wheezing (31%) and atopy (38%).
- We selected **N=818 children** with complete immunological data, i.e valid measures of the following immune markers:
  - **Four cytokines (IFN-γ, IL-5, IL-13, and IL-10)** obtained under five different immunological conditions (spontaneous response, mitogen response, three different antigen specific response (*ascaris l., b. tropicalis, dermatophagoides pter.)*
  - ***Four different specific IgEs*** (sIGE) specific to antigens (*dermaphagoides pter., b. tropicalis, b. germanica, p. americana*)

Step 1-Grouping immune markers

Objective:

- Translate the existing immunological knowledge to a **conceptual model** graphically representing the proposed relationships of each immune marker to an underlying immunological mechanism and the interrelationship between these.
- **Group the immune markers according to the conceptual model**
- Often a **multilevel causal framework** is necessary integrating multiple markers at different levels.
- For example in allergy research, immunologists classify cytokines according to distinct types of immune responses (Th1 response, Th2 response, or regulatory (T-Reg) response).

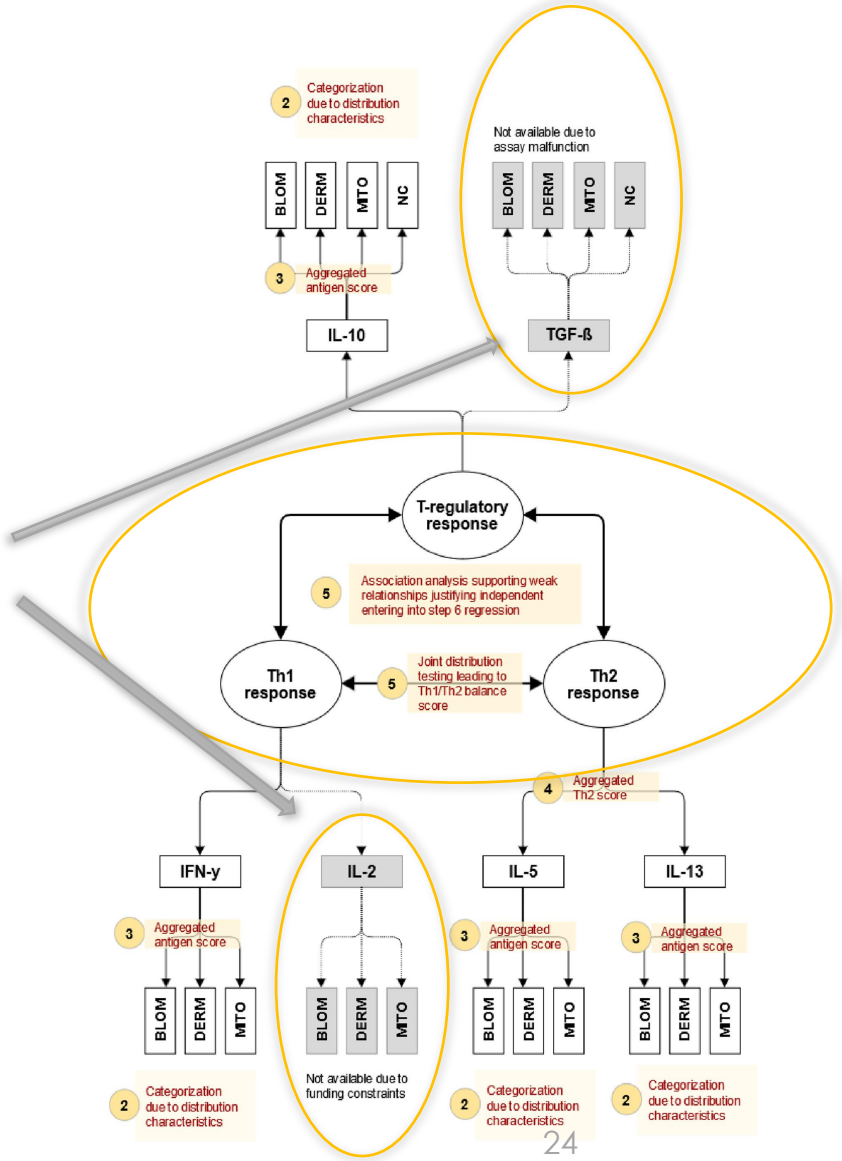
Conceptual framework for epidemiological analyis of atopic disease



bg stats

Conceptual model: SCAALA cytokine data

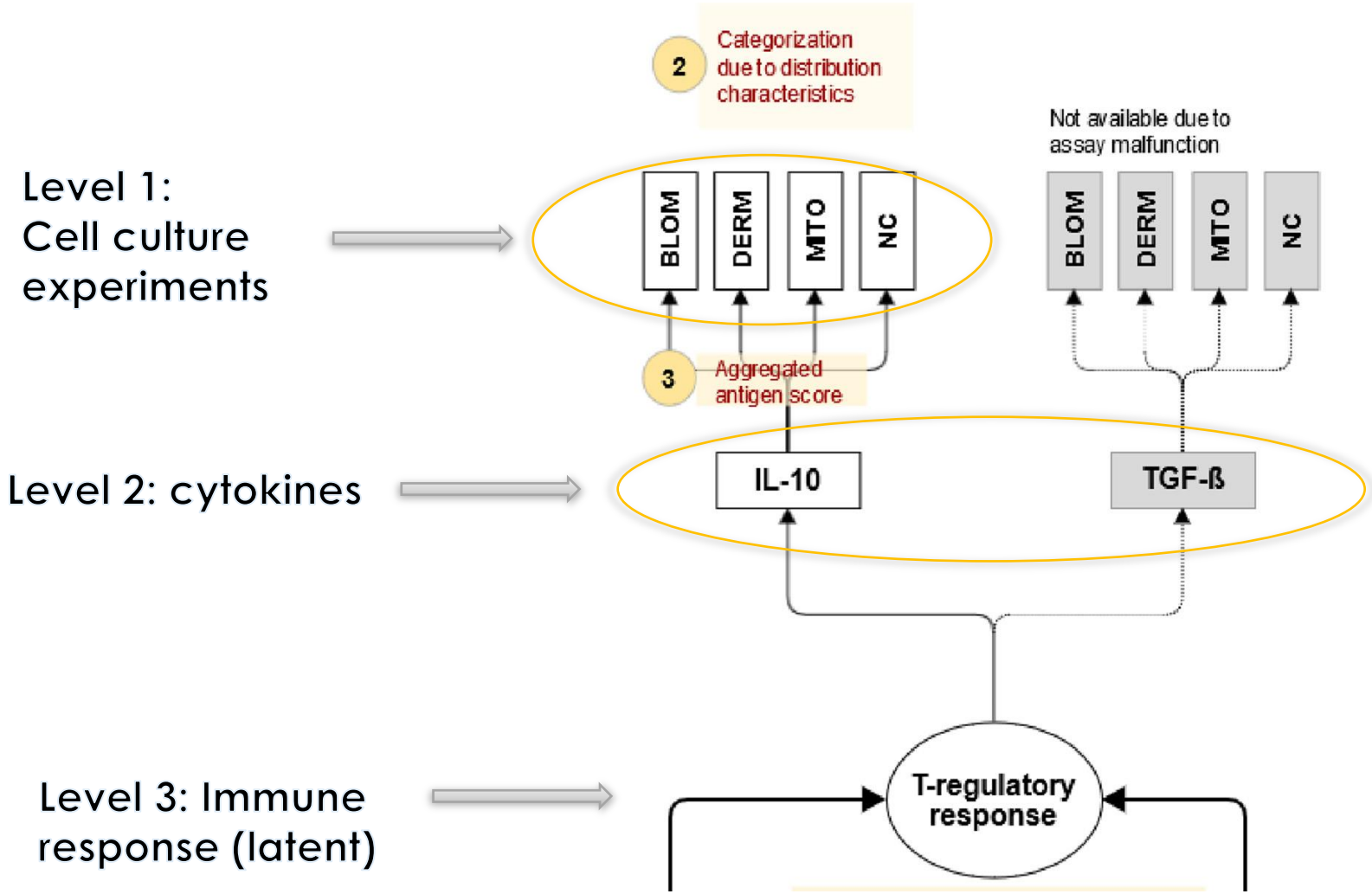
Data not available



Latent immunological concepts

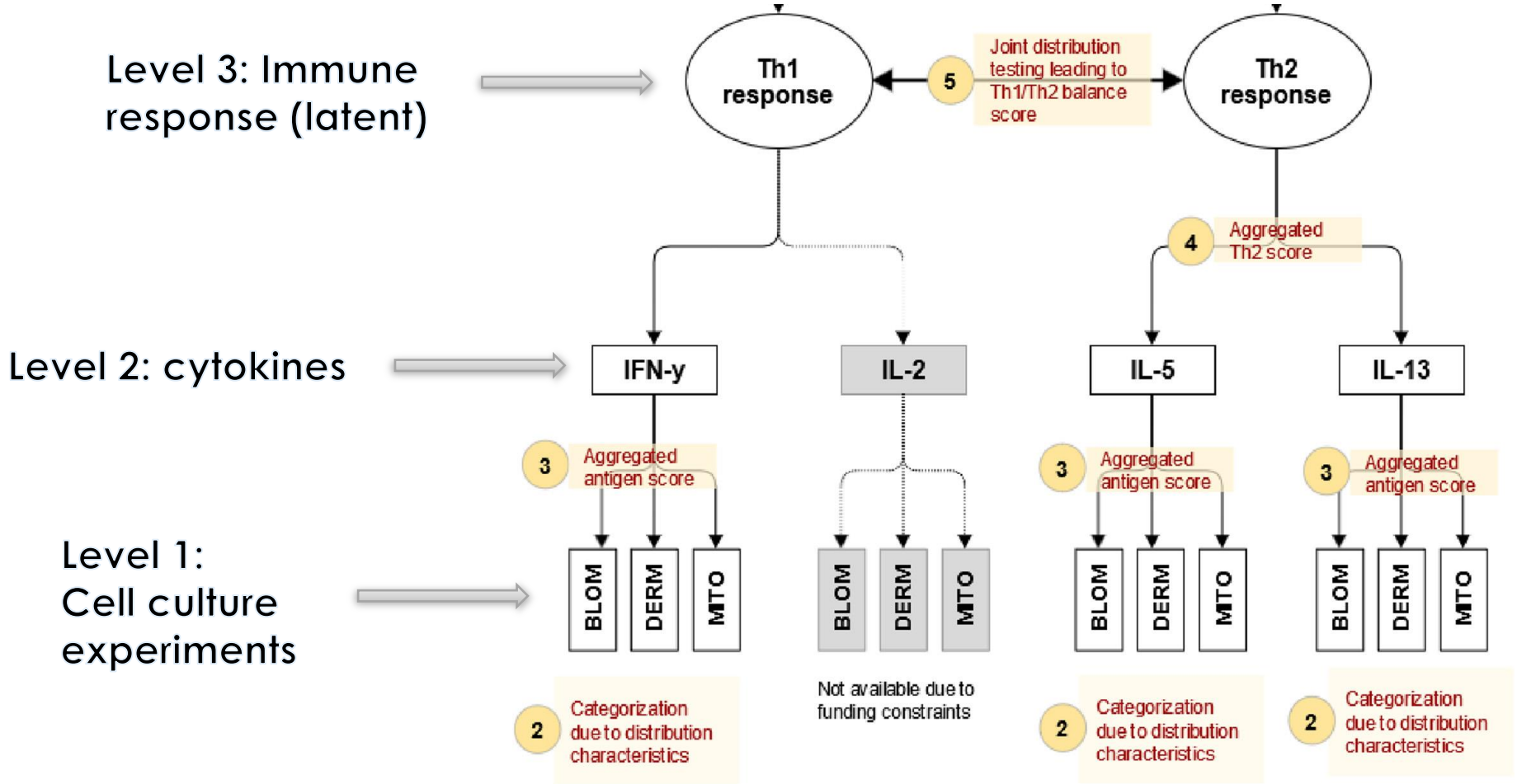
bg stats

Multi-level model for T-Reg response



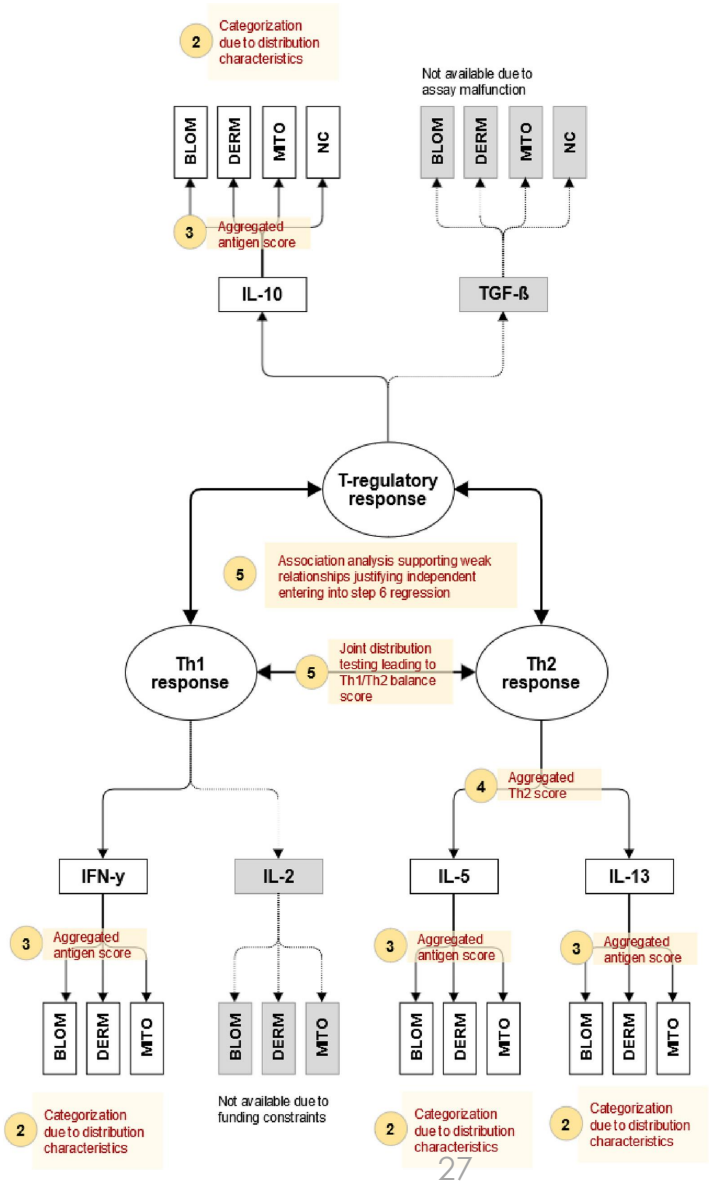
bg stats

Conceptual model part 2: Multilevel model for Th1- & Th2- response



bg stats

Conceptual model: SCAALA cytokine data



bg stats

Step 2 – Data exploration, data recoding

- To examine the statistical properties of each measurement and to employ statistically appropriate **transformations and recoding**

- If data are truncated (eg by "detection limits") special approaches for censored data such as Kaplan-Meier method or a generalized Wilcoxon test may be applied

- Eventually **recode data into distinct response categories** (e.g. no -, low – vs. high response)
  - Cut-points for defining response categories must be carefully chosen.
  - If „immunological cut-points" do not exist, categorisation should be based on characteristics of the distribution such as quantiles (e.g. median, tertiles, or quartiles).

Step 3: Intramarker (intragroup analysis)

- Objective: To study the **associative patterns within multiple markers** and statistically justified **aggregation to summary scores**.
- For example, association analysis of different stimulation methods for the same cytokine in culture experiments may suggest deriving one or more stimulation aggregate scores.
- Which method is appropriate for inter-dependence analysis depends on the properties of the measurement identified in step 2.
- Whether the data can be aggregated to scores depends on the pattern identified

Step 3: Methods for Intra-marker (intra-group analysis)

Table 1 Statistical approaches for interdependence analysis of immunological markers dependent on the scale of measurement

Scale of measurement	Bivariate methods	Multivariate methods
Binomial (e.g. positive, negative)	Contingency table; tests: Chi-square or Fisher's exact test; association measure: phi coefficient, Yule's Q	Multilayer contingency table, classification trees
Nominal (e.g. Th1, Th2, or T-Reg)	Contingency table; tests: chi-square or Fisher's exact test; association measure: contingency coefficient	Multilayer contingency tables, correspondence analysis, classification trees
Ordinal (e.g. low, medium, high)	Contingency table; tests: chi-square or Fisher's exact test, tau test; association measure: Spearman-Rank correlation, Kendall's Tau or Goodman and Kruskal's γ	Multilayer contingency tables, correspondence analysis, classification trees
Continuous (non-normal distributed)	Scatter Plots; test: Spearman-Rank Correlation, criteria: Kendall's Tau or Goodman and Kruskal's γ	Factor analytic techniques: e.g. principal component analysis
Continuous (normal distributed)	Scatter Plots; test and association measure: Pearson correlation coefficient	Factor analytic techniques: e.g. principal component analysis

Rules of thumb for quantifying the strength of association based on the magnitude of association measures (e.g. Goodman and Kruskal's γ): no association: $0 < |\gamma| <= 0.25$, weak: $0.25 < |\gamma| < 0.50$, moderate: $0.50 < |\gamma| <= 0.75$, strong: $|\gamma| > 0.75$

Step 3: Data aggregation to intra-marker scores

Potential aggregation depends on the intra-marker pattern identified:

1) Strong but simple positive association patterns indicate the **presence of a common major underlying immunological mechanism:** Combine measurements using simple aggregation functions, such as *the average response* (for continuous data) or *the maximum response* (for categorical or ordinal data).

2) Weak to moderate associations (or a *more complex association) pattern indicate the presence of several in part related (overlapping) immune mechanisms:* Reducing the multidimensional data to an immunologically meaningful factor solution (eg by PCA or CA). The resulting summary scores are weighted averages of the original markers, with factor loadings representing the weights.

3) No *meaningful pattern indicates* *that the* measures are "immunologically independent", likely to represent distinct underlying immunological phenomena. The aggregation of such independent markers to simple summary scores (e.g. the *sum of response)* should be avoided because information might be lost and interpretation of summary variables complicated ("apples + pears").

Step 4: Intermarker analysis

- If data are conceptually structured in multiple levels, e.g. an immunological mechanism (level 3) being operationalized by responses of different cytokines (level 2), which in turn were obtained from measurements during multiple stimulation assays (level 1), the **analytical strategy of step 3 should be repeated on the obtained aggregate scores from step 3**.
- Again the objective is to **explore association patterns** allowing to further aggregate the data to immunological summary scores.

For example, immunologists assume that both **IL-5 and IL-13** are produced by **Th2 cells** and thus show a strong association pattern due to this common mechanism (Th2 related immune response). Aggregating the antigen-specific summary scores for IL-5 and IL-13 (derived from step 3) to a **Th2 summary score** may be justified.

Step 5: (Inter) dependence analysis among immunological summary scores

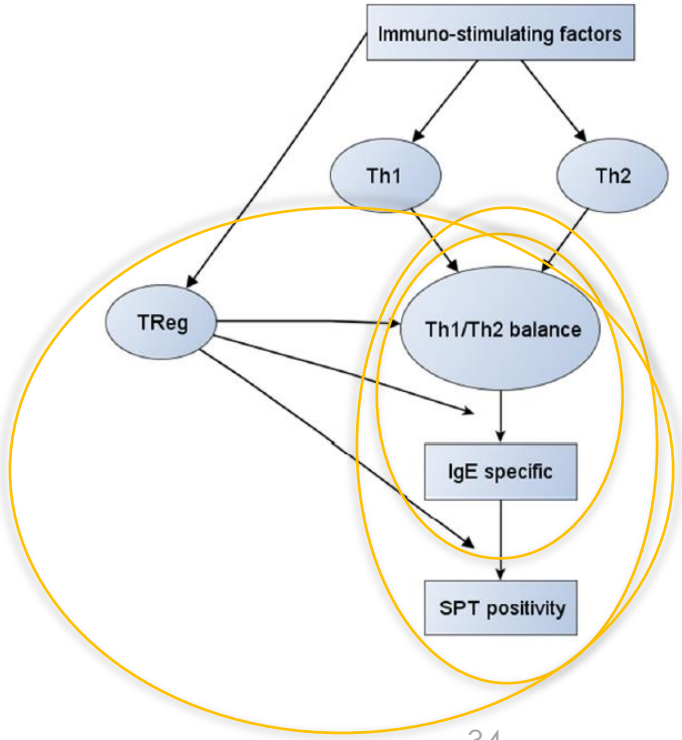
- **Interdependence analysis:** If there is no clear a priori hypothesis with respect to the **direction of the relationship**, the analytical approaches from steps 3 and 4 are repeated using now the summary scores.

- **Dependence analysis:** If we assume an underlying mechanism, the assumptions about the direction of causality should be incorporated in the model.

  - For example, if we assume that T-Reg responses affect the Th2 response (and not vice versa), a regression model using the Th2 score as the dependent variable and T-Reg score as the independent (predictor) variable may be warranted and preferred to a simple correlation analysis

Step 6: Dependence analysis based on immunological summary scores

Relating **the immunological summary scores** derived in steps 1 to 5 **to outcomes typically observed in immunological studies** (e.g. clinical outcomes like skin prick test, asthma symptoms, or intermediate immune markers such as plasma sIgE concentrations):



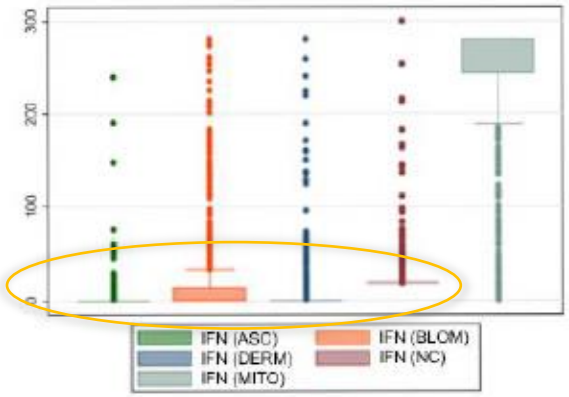
bg stats

RESULTS

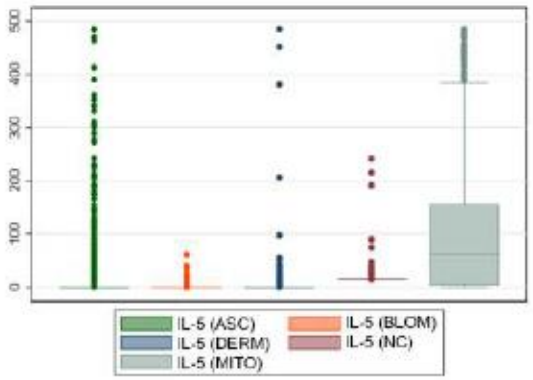
Results step 2a: Data exploration: Original cytokine concentrations

Truncation:
Many observations below the detection limit!

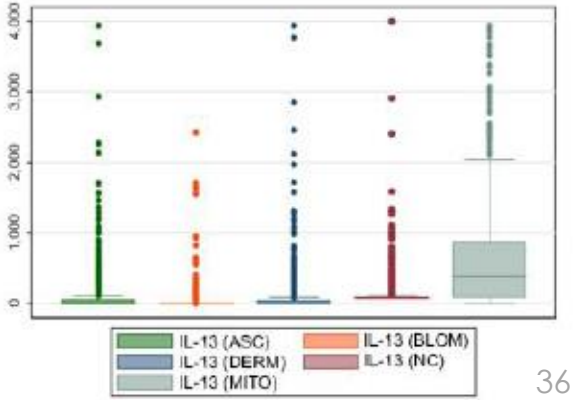
A) IFN-γ:



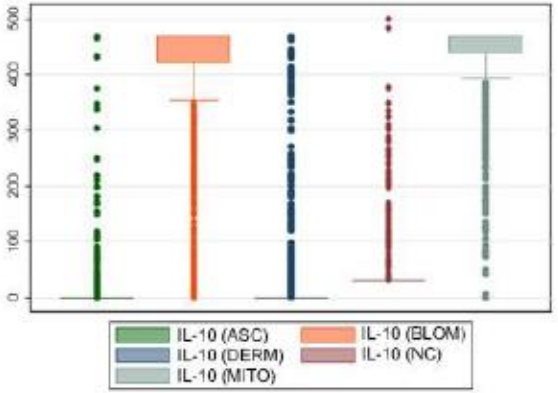
B) IL-5:



C) IL-13:



D) IL-10:



bg stats

Results step 2b: Recoding the cytokine data

1) Recoding Non-detects:

- Cytokine measurements below detection limits were considered "non-responders" and assigned a cytokine concentration equal to the lower detection limit
- Measurements above the detection limit were assigned a concentration equal to the upper detection limit.

2) Recoding continuous data to ordinal data:

- Children with detectable levels of cytokines (i.e. responders) were assigned to ordered categories with **cut-offs based on median or tertiles** for each cytokine and culture condition.
- IL-5 and IL-13 responses to mitogen, for which there was a higher proportion of responders, were classified into four categories (non-, low, intermediate, and high responders).
- All other measures were classified into three categories (non-, low, and high responders).

Results step 2b: Data recoding

Cytokine	ASC		BLOM		DERM		MITO		NC	
IFN-γ	n	%	n	%	n	%	n	%	n	%
no response	796	97.3	620	75.8	757	92.5	71	8.7	728	89.0
low response	12	1.5	95	11.6	31	3.8	137	16.7	45	5.5
high response	10	1.2	103	12.6	30	3.7	610	74.6	45	5.5
IL-5	n	%	n	%	n	%	n	%	n	%
no response	712	87.0	810	99.0	801	97.9	183	22.4	765	93.5
low response	56	6.8	4	0.5	8	1.0	206	25.2	27	3.3
intermediate response	-	-	-	-	-	-	217	26.5	-	-
high response	50	6.1	4	0.5	9	1.1	212	25.9	26	3.2
IL-13	n	%	n	%	n	%	n	%	n	%
no response	658	80.4	767	93.8	666	81.4	140	17.1	532	65.0
low response	78	9.5	25	3.1	71	8.7	211	25.8	141	17.2
intermediate response	-	-	-	-	-	-	231	28.2	-	-
high response	82	10.0	26	3.2	81	9.9	236	28.9	145	17.7
IL-10	n	%	n	%	n	%	n	%	n	%
no response	780	95.4	24	2.9	628	76.8	21	2.6	749	91.6
low response	19	2.3	153	18.7	85	10.4	121	14.8	36	4.4
high response	19	2.3	641	78.4	105	12.8	676	82.6	33	4.0

bg stats

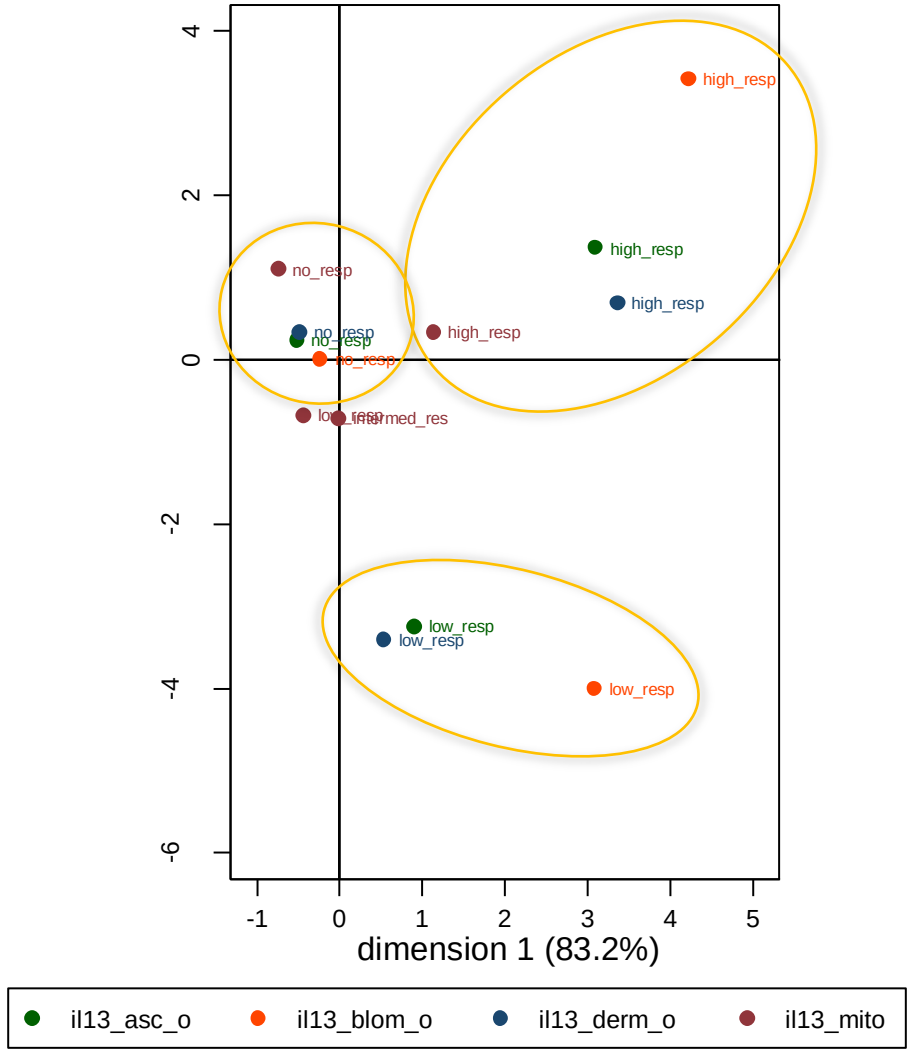
Results from step 3: Intra-marker analysis

Table: Summary of bivariate association analyses between each culture condition for each cytokine

Cytokine	Measure	ASC	BLOM	DERM	MITO
IFN-γ	BLOM	γ = 0.21, P = 0.144	-	-	-
	DERM	**γ = 0.77***, P = <0.001**	**γ = 0.73**, P = <0.001**	-	-
	MITO	γ = -0.24, P = 0.667	**γ = 0.26*, P = 0.016**	γ = 0.09, P = 0.653	-
	NC	**γ = 0.29*, P = 0.242**	γ = 0.05, P = 0.593	γ = 0.11, P = 0.412	**γ = -0.84***, P = <0.001**
IL-5	BLOM	γ = -0.05, P = 0.585	-	-	-
	DERM	**γ = 0.35*, P = 0.057**	**γ = 0.94***, P = <0.001**	-	-
	MITO	**γ = 0.26*, P = 0.047**	γ = -0.21, P = 0.234	γ = 0.19, P = 0.568	-
	NC	**γ = 0.38*, P = <0.001**	**γ = 0.61**, P = <0.001**	**γ = 0.76***, P = <0.001**	γ = 0.09, P = 0.405
IL-13	BLOM	**γ = 0.83***, P = <0.001**	-	-	-
	DERM	**γ = 0.83***, P = <0.001**	**γ = 0.85***, P = <0.001**	-	-
	MITO	**γ = 0.47*, P = <0.001**	**γ = 0.28*, P = 0.016**	γ = 0.52**, P = <0.001	-
	NC	**γ = 0.43*, P = <0.001**	**γ = 0.70**, P = <0.001**	**γ = 0.45*, P = <0.001**	**γ = 0.09, P = 0.005**
IL-10	BLOM	**γ = 0.28*, P = 0.510**	-	-	-
	DERM	**γ = 0.69**, P = <0.001**	**γ = 0.35*, P = <0.001**	-	-
	MITO	**γ = 0.26*, P = 0.326**	**γ = 0.93***, P = <0.001**	**γ = 0.43*, P = <0.001**	-
	NC	γ = -0.03, P = 0.468	**γ = -0.95***, P = <0.001**	**γ = -0.35*, P = 0.008**	**γ = -0.96***, P = <0.001**

bg stats

Results from step 3: Correspondence analysis IL-13



bg stats

Results from step 3: Data aggregation to cytokine specific scores for antigen response

Table: Distribution of aggregated summary scales for the antigen-specific response

Cytokine	IFN-γ		IL-5		IL-13		IL-10	
	n	%	n	%	n	%	n	%
no response	604	73.8	796	97.3	652	79.7	22	2.7
low response	94	11.5	11	1.3	77	9.4	152	18.6
high response	120	14.7	11	1.3	89	10.9	644	78.7
Total	818	100.0	818	99.9	818	100.0	818	100.0

bg stats

Results from step 4: Intermarker analysis

Table : Summary of results of inter-cytokine analysis using different scales (ANTI, MITO, NC).

Scale	Cytokine	IL-5	IL-13	IL-10
ANTI	IFN-γ	γ = -0.36*, P=0.669	γ = 0.12, P=0.324	**γ = 0.33*, P=0.011**
	IL-5		**γ = 0.58**, P=0.001**	γ = 0.06, P=0.242
	IL-13			γ = -0.04, P=0.648
MITO	IFN-γ	**γ = 0.53**, P<0.001**	**γ = 0.65**, P<0.001**	**γ = 0.46*, P<0.001**
	IL-5		**γ = 0.62**, P<0.001**	**γ = 0.27*, P<0.001**
	IL-13			**γ = 0.19, P=0.008**
NC	IFN-γ	γ = 0.18, P=445	γ = 0.09, P=0.095	γ = 0.04, P=0.500
	IL-5		γ = -0.22, P=0.451	γ = -0.39*, P=0.620
	IL-13			γ = 0.07, P=0.484

bg stats

Summary step 4: Intermarker analysis

- **Intercytokine analysis** examined the association patterns between any pair of cytokines on three different scales (ANTI, MITO, and NC).
- **Th2-related cytokines IL-5 and IL-13 showed positive associations** on both antigen and mitogen scales ($\gamma = 0.58$, $P = 0.001$ for ANTI, and $\gamma = 0.62$, $P < 0.001$ for MITO), but no association on the spontaneous scale ($\gamma = -0.22$, $p = 0.451$ for NC).
- We calculated a **Th2 summary score** by considering the highest observed response category of either Th2-related cytokine (**maximum response of IL-5 and IL-13**).
- We calculated the Th2 score for both the antigen and mitogen scale.
- In addition, we quantified the **balance of Th1 vs. Th2 cytokines** by considering the joint distribution of Th1-and Th2 responses.

Results from step 4: Immunological summary scores

Table : Distribution of final immunological summary scores

T-Reg response	ANTI		MITO		NC	
	n	%	n	%	n	%
no response	22	2.7	21	2.6	749	91.6
low response	152	18.6	121	14.8	36	4.4
high response	644	78.7	676	82.6	33	4.0
Th1 response	ANTI		MITO			
	n	%	n	%		
no response	604	73.8	71	8.7		
low response	94	11.5	137	16.7		
high response	120	14.7	610	74.6		
Th2 response	ANTI		MITO			
	n	%	n	%		
no response	642	78.5	99	12.1		
low response	79	9.7	162	19.8		
intermediate response	-	-	218	26.7		
high response	97	11.8	339	41.4		
TH1/TH2 balance	ANTI		MITO			
	n	%	n	%		
no TH1 resp. / no TH2 resp.	481	58.8	36	4.4		
TH1 resp./no. TH2 resp.	161	19.7	63	7.7		
TH1 resp./TH2 resp.	53	6.5	684	83.6		
no TH1 resp./TH2 resp.	123	15.0	35	4.3		

bg stats

Results from step 5: Inter-dependence analysis

Table : Results of interdependence analysis among the immunological summary scores

Immune response		Th2 response		T-Reg response		
	Scale	ANTI	MITO	ANTI	MITO	NC
Th1 response	ANTI	γ = 0.06, P=0.248	**γ = 0.25*, P<0.001**	**γ = 0.33*, P=0.011**	γ = 0.14, P=0.504	γ = 0.11, P=0.302
	MITO	γ = -0.01, P=0.723	**γ = 0.59**, P<0.001**	**γ = 0.42*, P<0.001**	**γ = 0.46*, P<0.001**	**γ = 0.26*, P=0.021**
Th2 response	ANTI	-	-	γ = -0.02, P=0.509	γ = -0.13, P=0.458	**γ = 0.30*, P=0.008**
	MITO	-	-	**γ = 0.36*, P<0.001**	**γ = 0.26*, P<0.001**	γ = 0.10, P=0.280
Th1/Th2 balance	ANTI	-	-	**γ= 0.14, P=0.030**	γ = -0.01, P=0.437	γ = 0.20, P=0.387
	MITO	-	-	**γ = 0.39*, P<0.001**	**γ= 0.17, P<0.001**	γ = 0.22, P=0.313

bg stats

Results from step 5: Inter-dependence analysis among the summary scores

Three important findings have emerged:

1) **No association between Th1 and Th2 (antigen)**

2) Th1/Th2 balance antigen showed to be **independent of T-Reg** (both spontaneous and mitogen)

3) A weak **association between Th2 response and T-Reg** on the antigen scale.

4) For **mitogen responses** all four summary scores showed **positive associations**.

Results from step 6: Dependence analysis with outcomes and immunological summary scores

Epidemiological association testing:

Outcome: **specific IgE max** (max response to any antigen).

Predictors: raw cytokine data or immunological summary variables

Confounders: age, gender

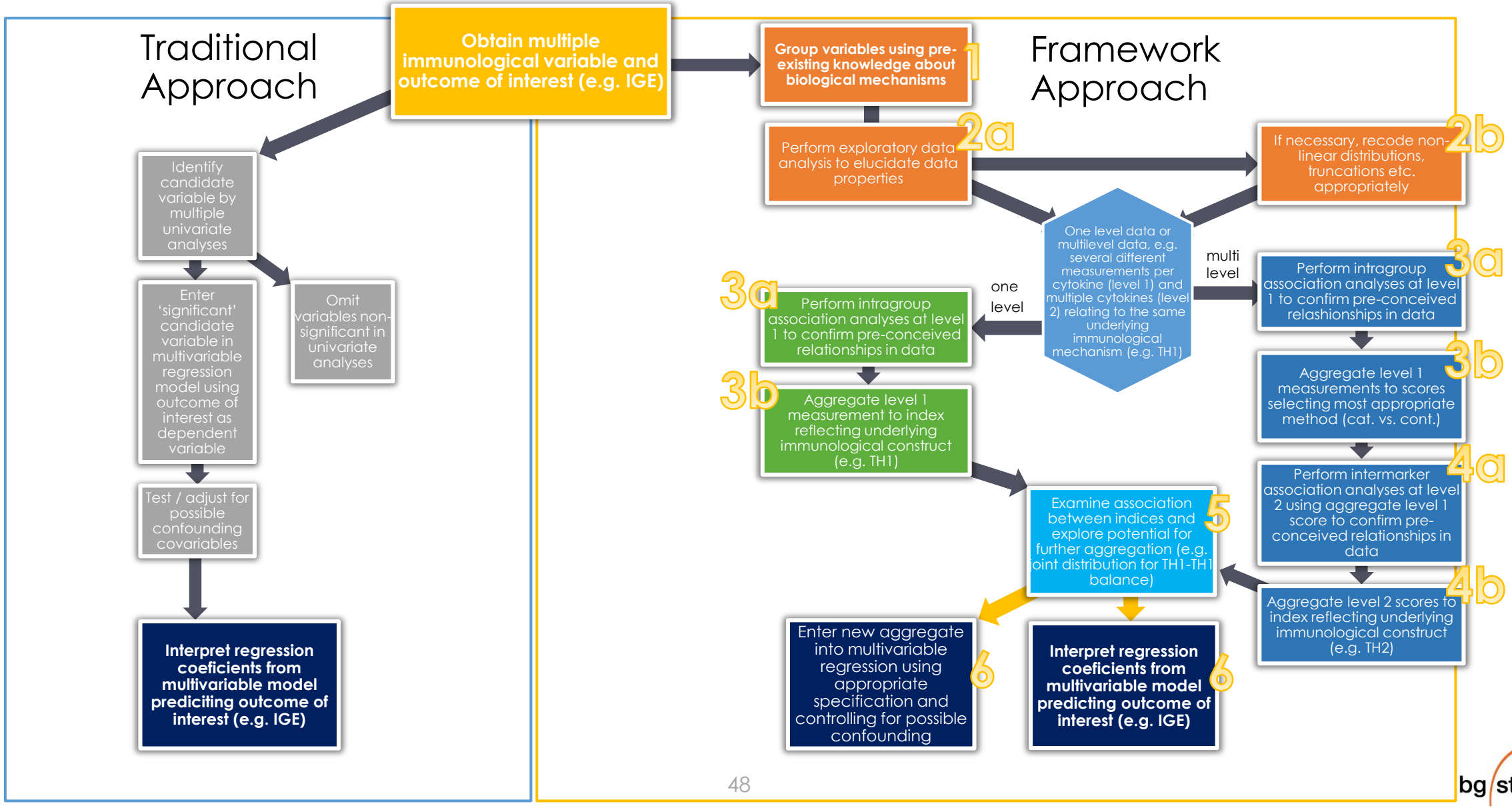
Model: linear regression

- As a consequence of the results of step 5 and our experts knowledge about the underlying mechanisms ***("extended hygiene hypothesis")*** we consider two independent mechanisms **(Th1/Th2 balance & T-Reg)** as independent immunological predictor variables.

- We consider the **antigen scale as the primary measure to quantify Th1/Th2** balance and the **spontaneous scale of T-Reg** to quantify the **immune regulatory mechanism.**

- We **compare** the results of the **framework approach** with a **classical stepwise regression approach** using the original markers (log-transformed raw cytokine data).

Comparison of Approaches



bg stats

Results from step 6: dependence analysis based on the immunological summary scores

Table : Results of traditional regression approach (final model after stepwise elimination

Parameter	Coef.	Std. Error	t	P>t	[95% Conf.	Interval]
IL-10 (DERM)	.0012180	.0004300	2.83	**0.005**	.0003739	.0020620
IL-5 (MITO)	.0012063	.0004971	2.43	**0.015**	.0002307	.0021820
IL-5 (BLOM)	.0674111	.0208156	3.24	**0.001**	.0265524	.1082699
IFN-γ (BLOM)	-.0037125	.0009908	-3.75	**0.000**	-.0056573	-.0017677
IL-5 (DERM)	.0067589	.0029443	2.30	**0.022**	.0009796	.0125382
Constant	-.6892018	.1016693	-6.78	**0.000**	-.8887674	-.4896361

Legend table A3: Regression coefficients reflect the change in log-transformed sIgE (unit: KU/l) per change in cytokine concentration (unit: pg/mL).

Table : Results of framework approach

Parameter	Coef.	Std. Error	t	P>t	[95% Conf.	Interval]
Th1/Th2 balance (ANTI)						
Th1+/Th2+	.3032749	.2946546	1.03	0.304	-.2750976	.8816473
Th1-/Th2+	**.5334773**	**.2029847**	**2.63**	**0.009**	**.1350422**	**.9319124**
T-Reg (NC)	-.3817676	.3668745	-1.04	0.298	-1.1018990	.3383640
Constant	-.5750811	.082135	-7.00	**0.000**	-.7363024	-.4138598

Legend: Regression coefficients reflect the change in log-transformed sIgE (unit: KU/l) per change in cytokine concentration (unit: pg/mL).

bg stats

CONCLUSIONS

Conclusions - Methodology

- We propose a **systematic analytical approach** for analysis of multiple correlated immune markers that **capitalizes on a conceptual framework** specifying the investigators' hypothesis about the underlying immunological phenomena.
- By **step-by step aggregating the information** from multiple correlated markers to **summary scores** our approach mimics the method of **latent variables modelling**.
- The **stepwise implementation is less data-driven** than classical latent variable approaches such as principal components analysis, latent class analysis or structural equation modelling because each analytical step is guided by a conceptual model.
- The resulting **non-redundant summary variables better reflect underlying immunological concepts** than the original markers and can be used in epidemiological **analysis** to quantify immunity.

Conclusions - Application

- The application of our approach to immune markers collected in SCAALA Salvador identified **three distinct immunological components**, Th1- response, Th2 - response and immune regulation that were only in part related:
  - On the antigen scale Th1 reponse and Th2 response showed to be independent components
  - In addition we considered the spontaneous T-reg response as component that was independent of both Th1 and Th2
- Association analysis with IgE levels showed findings in line with the extended hygiene hypothesis:
  - **Th2 skewness is a better predictor for elevated specific IgE levels than** Th2 response itself.
  - Further, we observed a non-significant tendency that **immune regulation downregulates IgE**.

    Caveat: Power was low! We need other studies with higher prevalence of strong immune regulation (high spontaneous IL-10) than SCAALA to confirm this potential !

Preview: Further applications

- The approach will be especially useful for **statistical data analysis involving multiple correlated immune markers** that are conceptually clustered due to the experimental designs and/or common causing underlying mechanisms
- Using **immunological summary scores** that reflect distinct immunological concepts instead of correlated original markers should substantially **simplify data analysis and enhance power** in studies on relationships among non-immunological factors, immune responses and disease.
- There is a large variety of **potential applications of the integrated approach** in modern immunology: Allergy research, inflammation(sepsis) research, infectious diseases, vaccine development, research on stress and immunity, etc.

"Let's focus on modelling immunological concepts rather than noisy immunological markers!!!"

Paper has been recently published …

Genser et al. BMC Immunology (2016) 17:11
DOI 10.1186/s12865-016-0149-9

BMC Immunology

METHODOLOGY ARTICLE Open Access

Applied immuno-epidemiological research: an approach for integrating existing knowledge into the statistical analysis of multiple immune markers

Bernd Genser[1,2*], Joachim E. Fischer[2], Camila A. Figueiredo[3], Neuza Alcântara-Neves[3], Mauricio L. Barreto[1,4], Philip J. Cooper[5,6], Leila D. Amorim[1,7], Marcus D. Saemann[8], Thomas Weichhart[9] and Laura C. Rodrigues[10]

bg stats

Acknowledgements

Funding:

- Scholarship from the *Brazilian Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (BR)*, contract no. 400011-2011-0 (Bernd Genser)
- Wellcome Trust Grant 088862/Z/09 (Phil Cooper)

Collaborating Academic Institutions:

- Instituto de Saúde Coletiva, Federal University of Bahia, Salvador, Brazil
- Instituto de Ciências da Saúde, Federal University of Bahia, Salvador, Brazil.
- Mannheim Institute of Public Health, Social & Preventive Medicine, University of Heidelberg, Germany
- London School of Hygiene & Tropical Medicine, UK
- Institute of Infection and Immunity, St George's University of London, UK
- Medical University of Vienna, Vienna, Austria.

bg stats

THANK YOU!

bernd.genser@bgstats.com
www.bgstats.com

bg stats