Analysing cluster randomized trials with count data by Frequentist and Bayesian methods. The BoliviaWET trial: Assessing the effect of SODIS on childhood diarrhoea

INAUGURALDISSERTATION
zur
Erlangung der Würde eines Doktors der Philosophie

vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Gonzalo Christian Durán Pacheco
aus
Potosí, Bolivia

Basel, February 2011

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. M. Tanner, Prof. Thomas Smith, Dr. Daniel Mäusezahl und Prof. Martin Röösli

Basel, den 26. Mai 2009

Prof. Dr. Eberhard Parlow
Dekan

This work is dedicated to Gabriela, José and Samuel my three reasons to live

Acknowledgments

My prime acknowledgements go to the three persons whose lives have been shaped in some ways during this PhD time.

To my wife, for her support from the moment we learned I was coming to Switzerland for my studies, but above all, for her love and inspiring fortitude during her *pursuit of happyness*. Gabriela, you were left on your own fighting battles beyond endurance day after day, as a mother, a father, a house provider and a first class professional. I learned from you the real meaning of resilience, to never giving up and to properly end whatever it is that I start. I am grateful for that, I admire you and I love you.

Josesito, they say we were like nail and finger, always together. All of the sudden you were left behind with a gap that your Mom has filled up so lovely. I missed you badly, I know you did too, I am ready now to catch up and recover our connection.

Samuelito, you were born when I was away in Switzerland running models, attending courses, being a PhD student. I missed your coming into this world and your first days of life. If I have something to regret it is not having been there next to your mom at that time.

Many people have contributed to the production of this dissertation. I wish to thank my supervisors Professor Thomas Smith and Dr Daniel Mäusezahl. Tom, for his significant inputs and his kind and friendly predisposition to sit down and guide me at all times. Thanks to Daniel, for his commitment toward my work and his valuable pieces of advice. Big thanks to Dr Jan Hattendorf my partner in many statistical undertakings. We set up a strong and fruitful collaboration, complementing statistical methodology and programming. Thanks Jan also for the Zusammenfassung.

At the STI, I am indebted to Dr Alan Shapira who kindly provided me access to his PC during weekends, holidays and his trips. Also to Bernadette Peterhans for the three laptops all together used to run the never ending simulations.

I thank Professor Marcel Tanner, for his warm reception, and for being the promoter of such a wonderful research institution called Swiss Tropical Institute. Special thanks to Christine Mensch, Christine Walliser and Margrit Slaoui, for their guidance and assistance in administrative matters. To my colleges at the office, from Claudia Sauerborn, Nafomon Segoba, Tippi Mak who already left, to Laura and Dominic Gosoniu, Nadine Riedel, Susan Rumisha, Bianca Plüss, Sandra Alba and Angelina Lutambi. Thanks for such a pleasant atmosphere.

The BoliviaWET trial is a realization of many people. I am grateful with them all for their warm welcome into the project and constant support. My thanks go to: Daniel Mäusezahl who set up the trial 6 years ago and our project leader; to Monica Daigl, Dr Markus Niggli, and Dr Jan Hattendorf who were my Swiss-side counterparts of data-managerial issues. I thank also Dr Andri Christen for those years of intercultural exchange during data collection in Bolivia and his kind hospitality from the very first moment I arrived in Switzerland. I wish to acknowledge the BoliviaWET field staff and my team members in particular, *the unit V* (data management unit) of BoliviaWET: Freddy Arauco a 100% reliable worker and my right hand during data entry supervision, Waldo, Edgar, Shirley, Edson and Angela, for their great work without any doubt a world class team. To our partners from the University of California at Berkeley, Dr Ben Arnold and Professor John Colford Jr, for the illuminating discussions on which much of the material in this dissertation was based on.

I am grateful to my family in law in Cochabamba, Don Willy, Doña Ruth, Mechi, Geral, Willy, Pablo, Tommy, Ana, for being there for us unconditionally during the good and bad moments during my absence. To my Mom Mela in Potosí and my Dad Pancho who has never gone. They always encouraged and supported me to achieve a university education. To my brother Ramiro in La Paz, for his friendship at the distance.

This research was funded through the National Institutes of Health (NIH), award number R01AI50087-01 and the Swiss Tropical Institute. I have been a recipient of a

stipend from the Stipendienkommission, Amt für Ausbildungsbeiträge of the Canton of Basel, Switzerland.

SUMMARY

Considerable attention has been given to the design and analysis of clinical trials where interventions are allocated to whole communities (e.g. schools, villages) rather than to individual participants. Such studies are known as cluster randomized trials or group randomized trials (CRTs). Motivated by the analysis of a community randomized trial (BoliviaWET) on solar water disinfection (SODIS) in Bolivia, this dissertation: *i)* outlines the primary analysis of the trial, *ii)* presents results from investigations undertaken to address analytical issues of situations observed in the trial and *iii)* presents results from topics of some secondary analysis. Statistical analysis was performed following both frequentist and Bayesian methods.

Chapter I gives a background on the established approaches for analysing CRTs. Some statistical methods are briefly described and the BoliviaWET trial is introduced. In addition, elements regarding the statistical analysis of BoliviaWET (e.g. design, model specification, selection of the statistical method) are discussed. The primary outcome, number of episodes per child per year was found to have substantial overdispersion. The Negative Binomial (NB) specification was found to satisfactorily address overdispersion. Generalized lineal mixed models were selected as the method for analysing the trial because of the reported overall good performance in analysing community randomized trial situations with small numbers of large clusters. Since the literature on the analysis of CRTs has mainly focused on binary and continuous data, a need for assessing methods for overdispersed counts was identified.

A full description of the trial and the main results are presented in chapter II. In summary, BoliviaWET was a CRT aimed at evaluating the effectiveness of SODIS to reduce diarrhoea among children under 5 in rural Bolivia. Twenty two rural communities participated in the study. The intervention, a comprehensive standardised SODIS promotion campaign, was randomly allocated to eleven communities following the pair-matched design. Diarrhoea occurrence of 376 children in the intervention arm and 349 children from the control arm was monitored for one year. Diarrhoea incidence was compared between arms producing an unadjusted (for covariates) relative rate of 0.81 (95% CI 0.59 - 1.12). The between-

cluster coefficient of variation CV_c was 0.27 (95% CI: 0.11 - 0.46). Parameters from the random-effect models were estimated via restricted pseudo-likelihood and MCMC on the basis of the considerations taken from chapter I. Results for adjusted models and analysis of other outcomes (prevalence, severe diarrhoea and dysentery) are also provided.

Chapter III studies the performance of five analytical methods for CRTs with overdispersed counts in settings similar to community randomized trials. The compared methods are: (i) The two-sample *t* test of cluster-level rates, (ii) Generalized estimating equations (GEE) with empirical covariance estimators (iii) GEE with model-based covariance estimators, (iv) Generalized Linear Mixed Models (GLMM) and (v) Bayesian Hierarchical Models (Bayes-HM). The NB distribution is applied to simulate overdispersed counts of CRTs with two study arms allowing the period of time under observation to vary among individuals. The effect of different sample sizes, degrees of clustering and degrees of cluster-size imbalance was investigated. The performance of the methods was assessed in terms of point, interval estimation and hypothesis testing properties.

Sample size and clustering led to differences between the methods in terms of CI's width, coverage, significance, power and random-effects estimation. GLMM and Bayes-HM performed better: Unbiased RR, nominal coverage, type I error rates and reasonable power. GEE showed higher power but anticonservative coverage and elevated type I error rates. The t-test yielded wide and unstable CI, the highest coverage and nominal significance. Imbalance affected the overall performance of the cluster-level t-test and the GEE's coverage in small samples. In explorations of the implications of ignoring overdispersion in the analysis of BoliviaWET data, upwardly biased RRs were observed for the Poisson analyses and the t-test. The existence of extreme values, more frequent in the control arm, violated the equidispersion assumption of Poisson analyses and the assumptions of the cluster-level t-test.

The point and interval estimation of the between-cluster coefficient of variation for overdispersed counts was studied in chapter IV. Four methods for point estimation were assessed: i) a cluster-level coefficient of variation (CL), ii) the CV_c from the one-way random-effect ANOVA, the root of the random-effect variance of iii) GLMM

and iv) Bayes-HM, both assuming NB distribution. The interval estimating methods were: i) Bootstrap confidence intervals (CI), ii) Generalized CI and iii) Bayesian credible intervals. Monte Carlo simulation was used to compare the methods at different sample sizes, and levels of clustering. The outcome was generated as NB counts with different individual period of follow-up.

GLMM and ANOVA both provided unbiased point estimates although ANOVA was more unstable under high clustering. CL heavily overestimated the between-cluster variation when it is lower or equal to 25%. Bayes-HM provided slight upward bias in settings without clustering. Bayes-HM performed best in terms of interval estimation. The effect of allowing for overdispersion was assessed by analysing the BoliviaWET dataset. Upwardly biased estimates were observed when assuming Poisson distribution. The magnitude of the bias resembled to that of the CL method observed in the simulations. The ANOVA-based approaches were not robust to the presence of extreme observations, being susceptible to producing anomalous random-effect estimates.

The meaning of the vernacular Quechua term *k'echalera* was evaluated as diagnosis of Diarrhoea in rural Bolivian settings (chapter V). Pre- and post-intervention data of BoliviaWET were employed where signs and symptoms of diarrhoea as well as *k'echalera* reports were recorded. Mother's reports of *k'echalera* were found to be associated with important changes in stool frequency, consistency and occurrence of blood and mucus. Interestingly, *k'echalera* reports were highly related to three types of watery-stool consistencies from the four applied in field tools. The *milky rice* stool consistency which fits into the definition of watery stool was not strongly related to *k'echalera*. Mucus in the stool was also associated with *k'echalera*. However its occurrence in *k'echalera*-free days accounted for at least 50% of the possible false negatives. Assuming an imperfect gold standard the sensitivity and specificity of the term *k'echalera* was estimated by Bayesian methods. We obtained a high specificity of at least 91% and sensitivity of at least 82% in average.

We investigated the factors that influenced on the adoption of SODIS in households in the intervention arm of BoliviaWET (chapter VI). Multivariable exploratory techniques were applied to identify typologies of SODIS users on the basis of 4

indicators of SODIS-use, and 2 indicators related to the duration of study participation. The chance of becoming a type of SODIS-user as a function of potential predictors was assessed by multinomial modelling. This subgroup analysis identified four groups of SODIS users after a 15-months extensive and comprehensive campaign. User-groups with high compliance were found to have a higher intensity of exposure to the SODIS campaign, latrine ownership, not having electricity, and having severely wasted children living in the home. The identified household factors related to the use of SODIS may help targeting populations that would benefit most from SODIS implementations. These findings indicate that pre-existing health knowledge, motivation and knowledge of disinfecting drinking water acquired through previous exposure to water, sanitation and hygiene programmes is associated with successful uptake of SODIS.

Finally, chapter VII provides a discussion of our main findings in context of the design of new cluster-unit trials and implications for statistical analysis, overdispersion and the methods applied in the secondary analysis,

In conclusion, the simulation studies suggest that GLMM and Bayesian models are appropriate for the analysis of overdispersed counts in CRTs in sample sizes ≤ 40 clusters in total. The estimation of the between-cluster coefficient of variation via GLMM and Bayes-HM is also appropriate. The Poisson model may seriously bias both the RR and CV_c estimates. The NB model with normal random-effects provides a natural way to address overdispersion of count data in a CRT. We, encourage to regularly verify the residual overdispersion and to apply the (Poisson or extra-Poisson) model that best fits the data.

The BoliviaWET trial found no strong evidence of reduction of the diarrhoea incidence in children <5 years in families using SODIS. In terms of secondary analyses, we conclude that the vernacular term *k'echalera* does refer to a change in the regular stool patterns associated with diarrhoea, although it differs from the symptoms-based diarrhoea definition in some aspects. We found that intensity of exposure to the SODIS campaign, latrine ownership, lack of electricity, and having severely wasted children living in the home are associated with the uptake of SODIS.

ZUSAMMENFASSUNG

Design-, Studienaufbau- und Analyseaspekte von klinischen Studien, bei denen die Randomisierungseinheit der Intervention nicht das Individuum darstellt, sondern ein Cluster von Probanden, wie z.B. Schulen oder Gemeinden, werden z.Z. wieder vermehrt diskutiert. Solche Studien werden als Cluster-randomisierte Studien (CRSs) oder Gruppen-randomisierte Studien bezeichnet. Motiviert durch die statistische Auswertung einer Cluster-randomisierte Studie zur Wirksamkeit von solarer Trinkwasserdesinfektion (SODIS) in Bolivien (BoliviaWET), umfasst diese Dissertation einerseits die Primäranalyse der Studie, andererseits Ergebnisse von Simulationsstudien zu speziellen analytischen Aspekten unter den in der Studie festgestellten Rahmenbedingungen und schliesslich die Resultate von weiterführenden Auswertungen. Die statistischen Analysen wurden dabei sowohl mit frequentistischen als auch mit Bayes'schen Methoden durchgeführt.

In Kapitel I werden einige Grundlagen zu den gängigen Analyseansätzen für CRSs beschrieben. Einige statistische Methoden werden kurz beschrieben und die BoliviaWET Studie wird vorgestellt. Zusätzlich werden einige Aspekte hinsichtlich der statistischen Auswertung der BoliviaWET Studie – wie Design, Modellspezifikation und Auswahl des statistischen Verfahrens – diskutiert. Es stellte sich heraus, dass die primäre Zielgrösse – Anzahl Durchfall-Episoden pro Jahr und Kind – eine substanzielle Überdispersion aufwies. Diese Streuung der Daten wurde durch Verwendung der negativen Binomialverteilung (NB) bei den Analysen angemessen berücksichtigt. Verallgemeinerte lineare gemischte Modelle (GLMM) wurden zur Analyse der Studie gewählt, da über generell gute Performance-Eigenschaften bei der Analyse von Studien mit einer geringen Anzahl, aber dafür relativ grossen Clustern, berichtet wurde. Die vorhandene Literatur zur Analyse von CRSs konzentriert sich hauptsächlich auf binäre und kontinuierliche Daten; eine kritische Beurteilung der Methoden im Zusammenhang mit Zähldaten ist in der Literatur bisher nicht verfügbar.

Eine detaillierte Beschreibung der Studie und die wichtigsten Ergebnisse werden in Kapitel II präsentiert. BoliviaWET war eine CRS um die Wirksamkeit von SODIS zur

Reduktion von Durchfällen bei Kindern unter fünf Jahren in ländlichen Gebieten Boliviens zu beurteilen. Zweiundzwanzig ländliche Gemeinden nahmen an der Studie teil. Elf Gemeinden wurden zufällig mittels Matched-Pairs Technik der Interventionsgruppe, bestehend aus einer intensiven und standardisierte SODIS-Werbe- und Schulungskampange, zugeteilt. Das Auftreten von Durchfall wurde bei 376 Kindern in der Interventionsgruppe und bei 349 Kindern in der Kontrollgruppe ein Jahr lang beobachtet. Die relative Rate (RR) der Durchfallinzidenz betrug 0.81 (95% CI 0.59 - 1.12) aus jenem Modell, welches einzig den Interventionseffekt und die Designfaktoren berücksichtigte. Der zwischen-Cluster Variationskoeffizient CV_c betrug 0.27 (95% CI: 0.11 - 0.46). Aufgrund der Überlegungen in Kapitel I, wurden die Parameter des Models mit zufälligen Effekten anhand der eingeschränkten pseudo-Maximum-Likelihood Methode ermittelt. Die Ergebnisse der adjustierten Modelle und der sekundären Zielkriterien (Prävalenz, Dysenterie, schwerer Durchfall) werden gleichfalls präsentiert.

In Kapitel III werden Leistungsindikatoren von fünf Analysemethoden, welche für die Auswertung von CRSs mit Überdispersion geeignet sind, unter Bedingungen getestet, die bei randomisierten Interventionsstudien mit Gemeinden als Cluster üblich sind. Die verglichenen Methoden waren: (i) Der Zweistichproben T-Test für Raten auf Clusterebene, (ii) verallgemeinerte Schätzgleichungen (Generalized Estimating Equations, GEE) mit empirischem Kovarianz Schätzer (iii) GEE mit Modell-bezogenem Kovarianz Schätzer, (iv) GLMM und (v) Bayes'sche hierarchische Modelle (Bayes-HM). In Simulationen wurden NB-verteilte Zähldaten mit Überdispersion generiert, wobei die Beobachtungsperiode individuell variierte. Untersucht wurde der Einfluss der Stichprobengrösse, Grad der Verklumpung (Clustering) und die Unausgewogenheit der Anzahl Probanden innerhalb der Cluster. Die Leistung wurde anhand der Güte von Punkt- und Intervallschätzer sowie Signifikanztests beurteilt.

Stichprobengrösse und Clustering führten zu Unterschieden bei den Methoden bezüglich der Weite des Konfidenzintervalls, Erfassungswahrscheinlichkeit des wahren Populationsparameters, Signifikanz, Power und Schätzung der zufälligen Effekte. GLMM und Bayes-HM erbrachten bessere Leistungen: unverzerrte RR sowie Erfassungswahrscheinlichkeit und Typ-I Fehlerraten nahe dem nominalem Niveau

und zudem eine angemessene Power. GEE war mit einer grösseren Power assoziiert, allerdings auf Kosten einer antikonservativen Erfassungswahrscheinlichkeit, die zu erhöhten Type-I Fehlerraten führte. Der T-Test lieferte weite und instabile Konfidenzintervalle, die höchste Erfassungswahrscheinlichkeit und eine Signifikanz nahe dem nominellen Niveau. Unausgewogene Clustergrössen beeinträchtigten die Performance von T-Test und GEE vor allem wenn die Stichprobengrösse klein war. Die Analyse der Daten der BoliviaWET Studie ergab, dass ein Missachten der Überdispersion bei Poisson verteilten Daten, und die Anwendung des T-Test bei NB-Verteilung zu aufwärts verzerrten RR führt. Extremwerte, welche in der Kontrollgruppe häufiger auftraten, verletzten die, bei der Poissonverteilung grundlegende Annahme der Equidispersion, sowie die Vorraussetzungen zur Durchführung des T-Tests.

Punkt- und Intervallschätzer des zwischen-Cluster Variationskoeffizienten (CV_c) für Zähldaten mit Überdispersion wurden in Kapitel IV untersucht. Dabei wurden vier Methoden zur Punktschätzung eingesetzt: i) ein Variationskoeffizient auf Cluster-Ebene (cluster level, CL), ii) der CV_c der einfaktoriellen Varianzanlyse mit zufälligen Effekten, i.e. die Quadratwurzel der Varianz der zufälligen Effekte, iii) GLMM und iv) Bayes-HM, beide mit NB Verteilung. Zudem wurden folgende Methoden der Intervallschätzung beurteilt: i) Bootstrap Konfidenzintervalle (CI), ii) verallgemeinerte CI und iii) Bayes'sche Intervalle. Anhand von Monte Carlo Simulationen wurden die Methoden bei verschiedenen Stichprobengrössen und unterschiedlichem Grad des Clusterings untersucht. Das Zielkriterium wurde als NB-verteilte Zähldaten generiert mit individuell variierendem Beobachtungszeitraum.

GLMM und die Varianzanalyse ergaben beide unverzerrte Punktschätzer, obwohl die Varianzanalyse bei starkem Clustering unstabilere Ergebnisse lieferte. CL überschätzte die zwischen-Cluster Varianz bei Werten kleiner oder gleich 25% stark. Bayes-HM erzeugte leicht erhöhte Resultate in Situationen ohne Clustering. Bayes-HM lieferte bei der Intervallschätzung das beste Ergebnis. Anhand der im Rahmen von BoliviaWET erhobenen Daten wurde der Einfluss von Überdispersion erörtert. Wenn das Zielkriterium als eine Poisson verteilte Variable analysiert wurde, war der CV_c generell zu hoch. Die Verzerrung war in etwa in der Grössenordnung der CL Methode während der Simulationen. Der varianzanalytische Ansatz war vor allem

anfällig gegenüber Extremwerten, wobei insbesondere das Schätzen der zufälligen Effekte negativ beeinflusst wurde.

Der Zusammenhang zwischen dem indigenen Qechua-sprachlichen Ausdruck *K'echalera* und der WHO Definition von Durchfall im ländlichen Bolivien wird in Kapitel V beschrieben. In einer Vorstudie wurden die Mütter zu verschiedenen Anzeichen und Symptomen von Durchfall, sowie dem Auftreten von *K'echalera* befragt. Das Auftreten von *K'echalera* war mit Veränderungen der Stuhlfrequenz und -konsistenz und blutigem oder schleimigen Stuhlgang assoziiert. Interessanterweise wurde *K'echalera* häufig in Kombination mit drei der vier flüssigen Stuhlkonsistenzkategorien des Fragebogens genannt. Dabei konnte bei der Kategorie "*milky rice*", welche ebenfalls eine flüssige Konsistenz beschreibt, kein Zusammenhang mit *K'echalera* festgestellt werden. Obwohl eine Assoziation zwischen schleimigen Stuhlgang und *K'echalera* bestand, war Schleim in über der Hälfte der Fälle vorhanden, bei denen die Symptome für eine Durchfallepisode nach WHO-Definition sprachen, aber *K'echalera* von den Müttern nicht genannt wurde. Unter der Annahme, dass die WHO Definition nicht als Goldstandard für Durchfall angesehen werden kann, wurden Sensitivität und Spezifität von *K'echalera* durch Bayes'sche Methoden bestimmt. Dabei wurden eine hohe Spezifität von mindestens 91% und eine Sensitivität von mindestens 82% festgestellt.

Mögliche Faktoren, die den Einsatz der SODIS Methode in der Zielbevölkerung der BoliviaWET Studie beeinflussen, werden in Kapitel VI beschrieben. Multivariable exploratorische Techniken wurden eingesetzt um die Haushalte in Nutzer-Klassen einzuteilen. Dafür wurden vier Indikatoren bezüglich der SODIS Applikation und zwei Indikatoren bezüglich der Länge der Teilnahme in der Studie herangezogen. Ein möglicher Einfluss verschiedener Faktoren wurde mit Hilfe von multinomialen Modellen. Diese Subgruppenanalyse identifizierte vier verschiedene Nutzergruppen die sich durch die 15-monatigen SODIS Kampagne gebildet hatten. Es stellte sich heraus, dass die Zugehörigkeit zu jener Gruppe, die SODIS am häufigsten praktizierte, durch einen intensivere Exposition zu der Intervention, das Vorhandensein einer Latrine, dem Fehlen von Elektrizität und stark ausgezehrter, schlecht ernährter Kinder begünstigt wurde. Die identifizierten Faktoren können helfen künftige Interventionen auf diejenigen Bevölkerungsgruppen auszurichten, die davon am meisten profitieren

können. Zudem signalisieren diese Ergebnisse, dass gesundheitsrelevantes Vorwissen, sowie bereits vorhandenes Wissen zur Trinkwasserreinigung aus der früheren Teilnahme an Wasser- und Siedlungshygiene Programmen zu einer beschleunigten Akzeptanz und Anwendung der SODIS Methode führt.

In Kapitel VII werden die Hauptergebnisse vor allem im Kontext von Aspekten des Studiendesigns diskutiert, die bei der Planung künftiger CRSs von Bedeutung sind. Zudem werden Empfehlungen für die statistische Methodenwahl bei der Primär- und Subgruppenanalyse und zum Vorgehen bei Überdispersion ausgesprochen.

Zusammenfassend kann festgehalten werden: die statistischen Simulationen zeigten, dass GLMM und Bayes'sche Modelle geeignet sind, um Cluster-oder Gemeinde-randomisierte Studien (CRS) mit Zähldaten und Überdispersion zu analysieren, selbst wenn die Anzahl der Cluster kleiner 40 ist. Auch beim Schätzen des zwischen-Cluster Variationskoeffizient (CV_c) zeigten GLMM und Bayes-HM gute Ergebnisse. Das Poisson Modell kann zu schwerwiegenden Verzerrungen sowohl beim Schätzen von RR als auch des CV_c führen. Das Negativ Binomial Modell unter Berücksichtigung von zufälligen Effekten stellt einen geeigneten Weg dar, um Zähldaten mit Überdispersion in CRSs zu analysieren. Es wird empfohlen, standardmässig die Überdispersion der Residuen zu verifizieren und das entsprechend beste Modell (Poisson oder extra-Poisson) zu wählen.

Die BoliviaWET Studie fand keinen stichhaltigen Nachweis für eine erhebliche Reduktion der Durchfallsinzidenz in Kindern unter fünf Jahren aufgrund einer SODIS Kampagne. Sekundäranalysen haben ergeben, dass der indigene Ausdruck *K'echalera* Änderungen des Stuhlgangs in der Form bezeichnet, wie sie für Durchfälle typisch sind. Trotzdem unterscheidet sich der Ausdruck in einigen Aspekten von der Standarddefinition der Weltgesundheitsorganisation. Eine intensive Exposition zur SODIS Intervention, das Vorhandensein einer Latrine, das Fehlen von Elektrizität und das Vorhandensein von schlecht ernährten und ausgezehrten Kindern im Haushalt waren mit einer erhöhten Akzeptanz und Anwendung der SODIS Methode assoziiert.

ABBREVIATIONS

ANOVA	Analysis of variance
Bayes-HM	Bayesian hierarchical models
BoliviaWET	Water evaluation trial in rural Bolivia
CI	Confidence interval | credible interval
CL	Cluster-level coefficient of variation
CRT	Cluster randomized trial
CS	Cluster specific
c.v.	Coefficient of variation
CV_c	Between-cluster coefficient of variation
EmpSE	Empirical standard error
GEE	Generalized estimating equations
GEE-Emp	Generalized estimating equations with empirical covariance estimates
GEE-MB	Generalized estimating equations with model based covariance estimates
GLM	Generalized linear models
GLMM	Generalized linear mixed models
GP	Generalized pivots
ICC	Intra class correlation coefficient
IQR	Inter quartile range
IR	Incidence Rate
MCMC	Markov chain Monte Carlo
NB	Negative Binomial
NGO	Nongovernmental organisation
NPV	Negative predictive value
OR	Odds ratio
PA	Population average
PCI	Project concern international
PET	Polyethyleneteraphtalate
POU-HWT	Point of use household water treatment
PPV	Positive predictive value
PR	Prevalence

RE	Random effects
RCT	Randomized controlled trial
cRCT	Cluster randomized controlled trial
RR	Relative rate
SAS	Statistical analysis system
SD	Severe diarrhoea
Se	Sensitivity
SODIS	Solar water disinfection
Sp	Specificity
T-test	Cluster-level two-sample *t* test
VIF	Variance inflating factor
WHO	World health organization

Table of Contents

CHAPTER I. Introduction

CHAPTER II. A cluster-randomized, controlled trial of solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia

CHAPTER III. Performance of analytical methods for overdispersed counts in cluster randomized trials: sample size, degree of clustering and imbalance

CHAPTER IV. Point and interval estimation of the between-cluster coefficient of variation for overdispersed counts in cluster randomized trials

CHAPTER V. Reporting diarrhoea through a vernacular term in Quechua speaking settings of rural Bolivia

CHAPTER VI. Factors associated with compliance among users of solar water disinfection in rural Bolivia

CHAPTER VI. Discussion and conclusive remarks

Appendices 177

LIST OF TABLES

LIST OF FIGURES

Chapter I:

Introduction

1. INTRODUCTION.

The allocation of health interventions in randomized controlled trials is often performed at the level of groups of individuals rather than the individual. These studies are known as group or cluster randomised trials (CRTs), and are considered the gold standard for the evaluation of health interventions when clusters (e.g. communities, hospitals, schools) are the units of random allocation. The case study examined throughout this thesis is a typical example of a field CRT. The intervention, the solar water disinfection method (SODIS), was randomly assigned to entire rural villages in Bolivia, while the outcome, childhood diarrhoea, was measured at individual level.

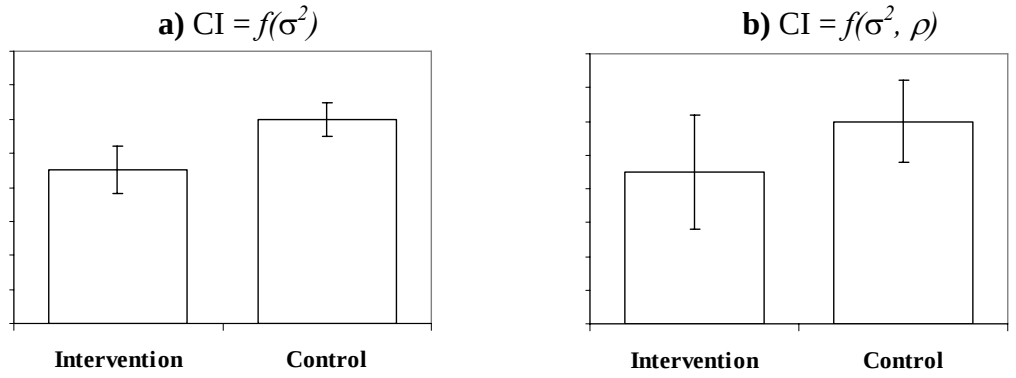
A distinctive feature of CRTs is that individuals from the same cluster are likely to respond in a more similar manner than units from different clusters, i.e. to have correlated responses. This potential violation of the independence assumption of standard statistical methods causes the underestimation of the true standard errors, leading to falsely narrow confidence intervals (CI) and fallaciously small *P values*. Indeed, if there is within cluster correlation, the variance of the outcome σ^2 becomes $VIF^*\sigma^2$, where:

$$VIF = 1 + (n-1)\rho$$

VIF denotes the *variance inflating factor* (*or design effect*) which depend on *n*, the number of individuals per cluster, and on ρ, the intra cluster correlation coefficient [1]. Figure I.1 illustrates the effect of clustering on the CI's width in a) a situation that ignores correlation and b) a situation that accounts for it.

That is why clustering must be allowed at the design and analysis stages, in order to avoid: i) elevated type 2 error rates for having underestimated the sample size to achieve a given power level, or ii) high type 1 error rates for having underestimated the standard errors during data analysis [2].

Figure I.1: Effect of ρ the intra cluster correlation coefficient on the width of CI of a two-arms CRT.



The methodological issues of CRTs have been broadly discussed in the statistics literature [2-4]. Specific topics of trial designs under a variety of practical conditions have been considered [5-10]. Similarly, analytical problems, assessment of statistical methods [11-22] and the need of effective reporting and proper interpretation has been also highlighted [23].

The remainder of this chapter provides introductory notes on analytical approaches for CRTs. This is followed by an introduction to the trial on solar water disinfection. A particular focus is given to design and analytical aspects of the trial, which will lead to the definition of the objectives of the thesis.

1.1. Overview to analytical methods for cluster randomized trials.

1.1.1. Analysis of cluster-level statistics

A straightforward way to address clustering during the analysis of CRTs is the use of cluster-level summary statistics. Individual-level data are combined within clusters to produce a cluster-level version of: the event rates, proportions, odds, means or the log versions of them. The intervention *versus* control analysis is performed by a t-test, a Wilcoxon's test, an ordinary least square regression or a meta-analysis random-effect regression of such summary statistics [12, 19, 24].

The principle of this type of analysis is the fact that the sample size considered for CI estimation of hypothesis testing is the number of clusters rather than the number of individuals. They are therefore conservative versions of the individual-level analysis where within-cluster variation is ignored. Some disadvantages are the obvious impossibility of adjusting for individual-level confounders in linear regression, potential bias [21], impossibility of assessing the within-cluster estimating precision and, related to the latter, a decrease in power and a loos of efficiency in trials with unequal cluster size [6].

1.1.2. Population average methods: GEE

Population averaged (PA) also known as marginal models, measure the effect of covariates on the mean response across the population, regardless of whether covariates vary within clusters [25]. Indeed, the mean response depends only on the covariates of interest and not on any (cluster) random effects, reflecting thus the average effect on the population.

PA methods separately model the mean response and the intra-cluster correlation. Correlation is considered a nuisance characteristic of the data when making inferences about the mean response. A well known method for fitting PA models is the generalized estimating equations (GEE). GEE is an extension of generalized linear models (GLM) where a link function is required to characterize the relationship of the mean response to a vector of covariates and a variance function to relate the variance of the outcome as a function of the mean [26]. Unlike GLM, no distributional assumptions are made in GEE and inferences are asymptotically unbiased and efficient as long as the mean and variance functions are correctly characterized. This method can be implemented in most major standard statistical packages [27] and is considered a natural approach to model the effects of interventions in CRTs because of the appealing interpretation of the marginal effects. There are however some technical problems regarding the use of the empirical (sandwich) variance estimator, when the number of clusters is lower than 50 [3].

1.1.3. Cluster specific methods: GLMM

In contrast to PA, the cluster specific methods (CS) are based on conditional models. Random effects are incorporated into the model to reflect correlation among observations made on the same cluster. CS account thus for an heterogeneity between subjects investigating and explaining the source of group to group variation, by modelling random effects along with fixed effect covariates.

Some examples are the random coefficient models, multilevel models, hierarchical regression, which can all be typified as a class of Generalized Linear Mixed Models (GLMM) [28]. GLMM represent an extension of GLM with a link and variance function specified along with the full distributional form of the response. The mean response is said to be conditioned on the (cluster) random effects and therefore they describe the cluster's response to changing covariates.

When CS contain covariates that do not vary within clusters the interpretation of the regression parameters can be complicated, because coefficients measure a contrast that is not observed in any single cluster [29]. A risk of underestimation of both fixed and random effects may occur when the level of clustering is large and the cluster size is small [3]. However, they have been reported to produce overall good performance when simulating situations similar to community-randomized trials [2]. The assumptions made on the random effects distributions are possibly the most important limitation. Misspecification of random-effects distribution may produce considerable bias both on the fixed effects coefficients and on their standard error estimates [30].

1.1.4. Bayesian analysis: hierarchical models

Bayesian methods are increasingly used in a variety of disciplines. They work with the notion of probability as a conditional measure of uncertainty, being the computation of posterior probabilities (probability of the parameters of interest given the data: $P(\theta \mid data)$ the focal concern. Empirical evidence from the collected data is combined with previous knowledge to produce such uncertainty measures, and a

posterior distribution of the parameters of interest is constructed by Markov chain Monte Carlo (MCMC) simulation [31].

In the framework of CRT, Bayesian methods deal with intracluster dependence in the same way as GLMM, by explicitly modeling the between-cluster variability through random effects. For instance, assuming a count outcome $Y \sim \text{Poisson}(\mu)$ of the intervention x_j (x = 0,1), the hierarchical model $\log(\mu_j) = \beta_0 + \beta x_j + \nu_j$ would reflect that the log of the expectations μ_j is a function of the intervention and the random effects ν_j of cluster j which follows a distribution with mean 0 and variance σ_c^2. The calculation of the posterior probabilities $P(\beta_0, \beta, \sigma_c^2 \mid Y)$ are done by updating the likelihood $f(Y \mid \beta_0, \beta, \sigma_c^2)$ with the prior $P(\beta_0, \beta, \sigma_c^2)$ as established by the Bayes' principle, through MCMC [13, 14].

Although the mathematical foundations of Bayesian methods are not discussed, the main point of controversy is the risk of incorporating subjectivity by the choice of prior beliefs. However 'non-informative' or 'reference' priors are widely used and it is also possible to investigate the sensitivity of the results to the priors [32].

2. ANALYSIS OF THE BOLIVIAWET TRIAL

2.1. Motivation for the trial.

Microbiologically safe water is considered an important determinant in preventing diarrhoeal disease in children under five years of age [33]. However, about 1.1 billion people lack access to improved water supplies [34]. Consequently several interventions (e.g. filtration, chlorination, boiling, flocculation) have been developed to improve water quality. Evidence showed that such interventions are in general effective in preventing diarrhoea, particularly when applicable at household level [35].

SOlar water **DIS**infection is a simple, low-cost and household water treatment method. It combines the effects of UV-A radiation and the increase of temperature in water exposed to sunlight in plastic bottles. Although SODIS has been proven to be efficacious at inactivating waterborne pathogens in laboratory conditions [36], there is not conclusive evidence of its health effects in populations without access to safe drinking water. Hence, a community randomized controlled trial was designed to assess the effectiveness of SODIS promotion in reducing diarrhoea among children under 5, without other access to clean drinking water.

2.2. Design.

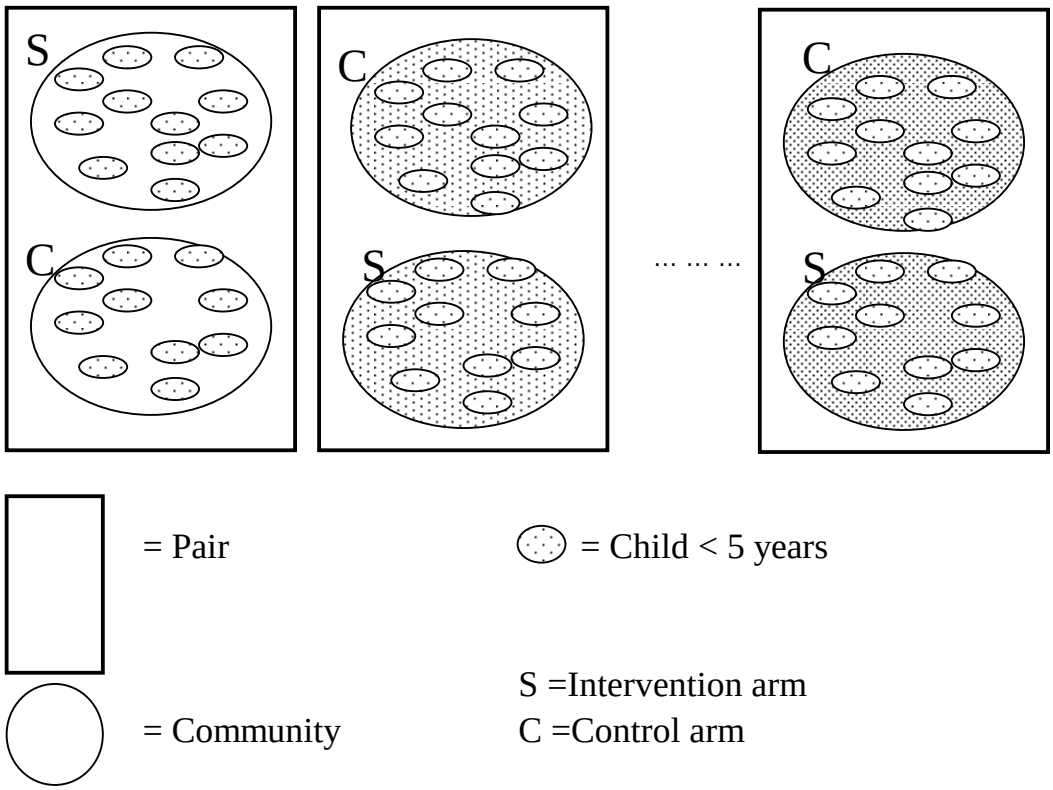
The intervention was a standardised interactive SODIS-promotion campaign. The study was designed with a rural village or community as the unit of random allocation. Reasons for cluster randomization are given in the design section of chapter II.

The trial design incorporated pair-matching and sample size was calculated allowing for clustering within communities by using methods proposed by Hayes & Bennett (1999) [5]. Sample size calculations suggested that at least 18 communities (9 pairs) with 10 persons-year of observation per community were sufficient to estimate a 33% difference, with a power of 80%, a significance of 0.05 and a between-cluster coefficient of variation (CV_c) = 0.20. Anticipating a drop-out of at least 2

communities and possible individual drop-outs, the final sample size was adjusted to 22 communities with 30 persons-year of observation.

The 22 rural villages were grouped in pairs by diarrhoea incidence as measured in an 8-week baseline survey. The SODIS intervention was then randomly allocated to one of the two communities within each pair (Figure I.2).

Figure I.2: Layout of the BoliviaWET study design



2.3. Primary outcome.

The primary outcome was the diarrhoea incidence defined as the number of diarrhoeal episodes per child per year at risk. In order to estimate the trial outcome, daily diarrhoea occurrence was monitored through a weakly health monitoring tool in 725 children from the 22 rural communities (detailed information is given in chapter II). Diarrhoea was measured as *K'echalera*, the local vernacular term (see chapter V). Additional related symptoms (frequency, consistency and presence of blood or mucus in the stool) were also collected.

The number of episodes for each child was calculated from the following definition. A new diarrhoeal episode was considered after at least 3 symptoms-free days [37, 38]. Table I.1 summarizes the incidence rates obtained in the two study arms as well as the observed relative rate (intervention over control).

Table I.1: Observed incidence rates in the two arms of the BoliviaWET trial

	Control	**Intervention**
Nr of children	349	367
Total Episodes	887	808
Children-days-at-risk	75077	82682
Group incidence rate	0.01181	0.00977
Crude Relative Rate (RR)	**0.827**	
Protective Effect (%)	**17.3**	

2.4. Statistical model.

Let us denote Y_{ijl} the number of episodes observed during t_{ijl} days at risk in the l^{th} child (l = 1,…,n_{ij}) from a community j allocated to an intervention group (j = 1,2) within pair i (i = 1,…,p). The statistical model for the pair-matched design above and specified in terms of generalized linear mixed models (GLMM) is the following:

$$\log(E[Y_{ijl}]) = \log(t_{ijl}) + \eta + B_i + \beta x_{ij} + \xi_{ij} \qquad (1)$$

i = 1,…,11 (Pair)

j = 1, 2 (communities allocated to the intervention group within the i^{th} pair)

l = 1,…,n_{ij} (nr of children from the j^{th} community from the i^{th} pair).

Where:

η = General log mean

B_i = Random effect of the i^{th} pair. $B_i \sim$ NIID[†] (0, σ_p^2)

β = the effect of the intervention, as the log-means (intervention-over-control) relative rate.

[†] NIID = Normally independent and identically distributed

x_{ij} = Intervention group (0 = control, 1=SODIS) allocated to the j^{th} community of the i^{th} pair.

ξ_{ij} = Random effect of the j^{th} community in the i^{th} pair. $\xi_{ij} \sim \text{NIID}(0, \sigma^2_{pc})$

The model implies a relative rate RR of $exp(\beta)$ and clustering accounted for through the random effects B_i and ξ_{ij} whose variances sum up to the total between-cluster variation $\sigma^2_c = \sigma^2_p + \sigma^2_{pc}$, i.e. between-pairs plus within-pairs variance, and ξ_{ij} used as an error term for testing $\beta = 0$.

2.5. Checking the model assumptions.

2.5.1. Examining residuals.

Two distributional assumptions were assessed for the outcome due to high overdispersion in the observed number of episodes per child and individual incidence rates (Table I.2):

Table I.2: Mean and variance of the nr of episodes per child and the individual incidence rates of the BoliviaWET Trial

	n	Mean	Variance
Nr of Episodes per child	725	2.3	8.6
Individual incidence rates*	725	5.5	269.6

*nr of episodes per child per year

i) $Y \sim$ Poison(μ) with variance function $V(Y) = \phi v(\mu) = \mu$ where ϕ the overdispersion parameter is assumed to be 1.

ii) $Y \sim$ Negative Binomial(s, μ) with a variance function $V(Y) = \phi v(\mu) = \phi\left(\mu + s\mu^2_{jl}\right)$, where ϕ is assumed to be 1 and s is the NB overdispersion parameter.

Results from the two analyses are summarized next (Table I.3). The residual overdispersion ϕ is clearly lower for NB compared to the Poisson model. Likewise, the information criteria (Pseudo AIC) is inflated for the Poisson model. This indicates

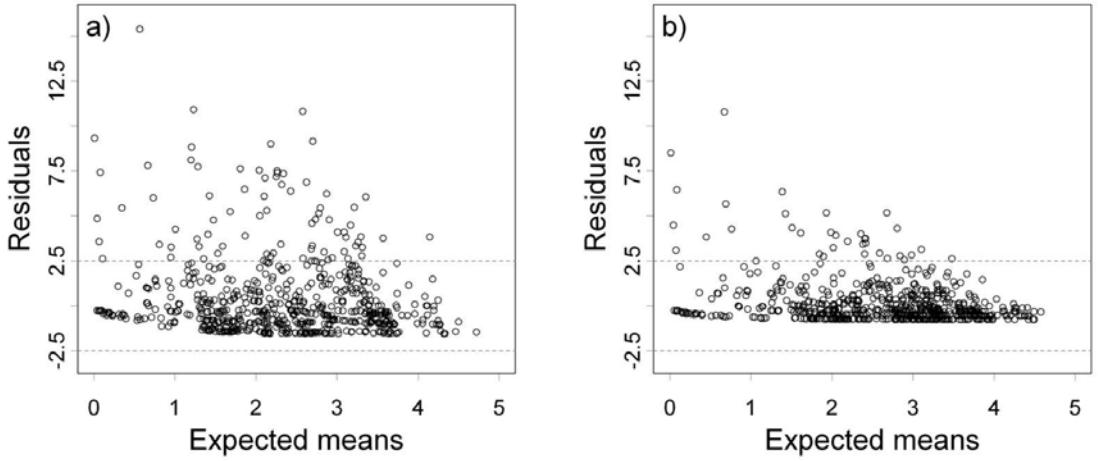
a clear better fit for the NB model. The analysis of the Pearson standardized residuals against the expected means confirms the better fit of the NB model (Figure I.3).

Table I.3: Comparison of the Poisson and Negative Binomial assumptions for model specification in the BoliviaWET data

		Poisson	**Neg Bin**
Fit statistics	*Pseudo AIC*[†]	4311.47	2769.62
	ϕ	4.74	1.28
Random effects	σ_p^2	0	0
	σ_{pc}^2	0.1049	0.07275
Fixed effects	β *(se)* [††]	-0.1421 (0.1473)	-0.2114 (0.1547)
	95% CI of β	(-0.4494, 0.1651)	(-0.5341, 0.1113)
	P-value	0.346	0.187

[†] Pseudo Akaike Information Criteria [††] se = Standard error

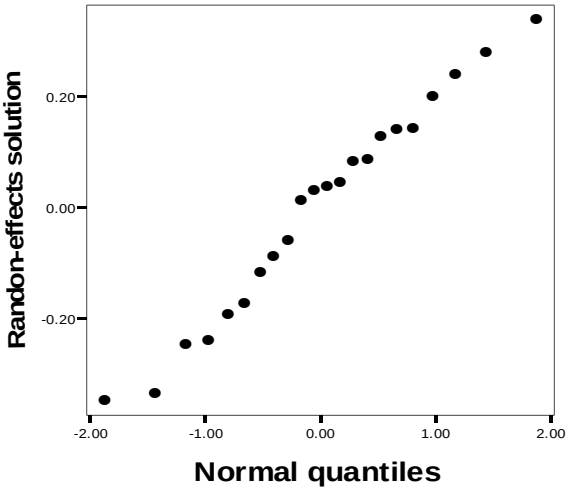
Figure I.3: Comparison of a) Poisson and b) Negative Binomial (Pearson-standardized) residuals of the BoliviaWET dataset.



2.5.2. Random-effects distribution.

The assumption of normality of random effects is difficult to assess in CRTs by statistical tests when the number clusters is small. Therefore, this assumption is checked using a normality probability plot [39]. Figure I.4 displays the random-effects predicted values against the expected values of the standard normal distribution. A straight line is indicative of normality. For the BoliviaWET data, correspondingly, the assumption of normally distributed random effects seems to be reasonable.

Figure I.4: Normal probability plot of the solution for random effects of the BoliviaWET trial.



2.6. Model selection.

Note that the between-pairs variance estimate in Table I.3 was 0 both for Poisson and NB analyses. This suggests that the between-cluster variance can obviate pair-matching because it was ineffective in controlling the outcome variance. Hence, model (1) can be reformulated to the random intercepts model, where a gain in power would be expected [2]. The MIXED and GLIMMIX procedures in SAS reformulate the model automatically when a variance component is found 0 [40, 41]. The fixed effects results in Table I.3 will be thus equivalent to the ones specified under the random intercepts model. In addition, and provided the better fit of the NB

distribution, the statistical model will be specified in terms of NB mean and variance functions (or the specification of the full distribution).

$$\log(E[Y_{jl}]) = \log(t_{jl}) + \eta + \beta x_j + \xi_j \qquad (2)$$

where $l = 1,\ldots,n_j$ (nr of children in the j^{th} community) $j = 1,\ldots,22$ (nr of communities), η the general log mean, β the change in the log-means (intervention-over-control) or log of the RR, x_j intervention group (0 = control, 1=SODIS) allocated to the j^{th} community, ξ_j the random effect of the j^{th} community ~ NIID$(0, \sigma_c^2)$.

2.7. Concluding remarks.

Based on the residual analysis we resolved for $Y_{jl} \sim$ NB(s, μ_{jl}) provided the better capacity of controlling the residual variance. The NB-random effects model may be considered a natural approach to account for overdispersion. It is equivalent to Poisson model with heterogeneous gamma-distributed means at individual level (within clusters) and normally-distributed cluster random effects.

The choice of GLMM over cluster-level or GEE methods relies upon the general support to GLMM in situations similar to community randomized trials, i.e. small number of large clusters [2, 3]. Previous literature reports GEE to underestimate the standard errors and to produce elevated type I error rates if the number of cluster is < 50 [2, 3]. On the other hand cluster-level methods may show low efficiency, elevated type 2 error rates and bias [21]. Note that such properties were observed in studies with continuous and binary data.

We present results by 5 statistical methods for CRTs (Table I.4) assuming the random intercepts model (2). This includes a Bayesian hierarchical regression, assuming $Y \sim$ NB(s, μ) with, uninformative priors: η ~ N(0, 10^6), $\beta \sim$ N(0, 10^6), $\sigma_c^2 \sim$ IG(0.001, 0.001), s ~ Γ(0.001,0.001) (IG=Inverse Gamma distribution). The exchangeable correlation structure is used for GEE.

Table I.4: Results from the analysis of the BoliviaWET data by methods for cluster randomized trials.

Parameter	Observed	T-test	GEE†	GLMM‡	NLMIXED*	Bayesian-HM
β	-	-	**-0.1707**	**-0.2114**	**-0.2042**	**-0.2154**
RR	**0.827**	**0.912**	**0.843**	**0.809**	**0.815**	**0.806**
(RR) 95% CI		(0.61, 1.20)	(0.64, 1.11)	(0.59, 1.12)	(0.59, 1.13)	(0.59, 1.10)
P-value		0.496	0.225	0.187	0.209	0.172

† using empirical variance estimator and exchangeable correlation structure (PROC GENMOD, SAS v1.9)

‡ GLMM with parameters estimated via Restricted Pseudo Likelihood (PROC GLIMMIX SAS v9.1.

* GLMM with parameters estimated via numerical integration (PROC NLMIXED, SAS v9.1).

The results in Table I.4 by the methods for CRTs merit a deeper attention. It is uncertain whether the properties of methods for continuous of binary data can be extrapolated to overdispered counts. Some studies have raised the issue of overdispersion and the comparison of methods when modelling count data [28, 29, 42, 43]. However, most of them were based on the analysis of real datasets where the true model parameters were unknown just like in Table I.4. Simulation studies are therefore required to assess the performance of methods for analyzing CRTs with overdispersed count data. Of additional importance is the need to identify appropriate approaches for estimating clustering under overdispersed count data situations of CRTs.

This dissertation is a synopsis of how such methodological and practical problems were dealt with during the primary and secondary analysis of the SODIS trial. Chapter II reports and discuss the main results of the trial. Chapter III provides findings of an evaluation made on the performance of analytical methods for CRTs applicable to overdispersed count data. Chapter IV similarly presents results from a simulation study on methods for point and interval estimation of the between-cluster coefficient of variation as the measure of clustering alternative to ρ. Chapter V reports the statistical validation of the local vernacular term used in the trial to account for diarrhoea in rural Bolivia. The analysis of factors associated to SODIS adoption in households that received the intervention is summarized in chapter VI. Finally an overall discussion of the main topics related to the design and analysis of CRTs in light of our experience, is presented in chapter VII.

3. OBJECTIVE OF THE THESIS

This work aimed at assessing methodological aspects of Frequentist and Bayesian analysis of overdispersed count data under typical situations of community randomized trials. In particular:

- To study the statistical performance (estimation and hypothesis testing) of analytical methods for CRTs with overdispersed count data, under situations analogous to real community intervention trials.

- To assess the performance of point- and interval estimating methods for the between-cluster coefficient of variation in situations analogous to real community intervention trials.

In addition, to contribute with analytical solutions to problems related to the secondary analysis of the trial such as:

- To validate the meaning of the vernacular term *k'echalera* to report child diarrhoea. in rural Bolivia

- To identify the factors that determine the adoption of SODIS

4. REFERENCES

1. Ukoumunne OC, Gulliford MC, Chinn S. A note on the use of the variance inflation factor for determining sample size in cluster randomized trials. *Journal of the Royal Statistical Society Series D-the Statistician* 2002; **51**:479-484.

2. Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials and Statistics in Medicine. *Statistics in Medicine* 2007; **26**(1):2-19.

3. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: A review of recent methodological developments. *American Journal of Public Health* 2004; **94**(3):423-432.

4. Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. *British Medical Journal* 2008; **336**(7649):876-880.

5. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *International Journal of Epidemiology* 1999; **28**(2):319-326.

6. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International Journal of Epidemiology* 2006; **35**(5):1292-1300.

7. Kerry SM, Bland JM. Unequal cluster sizes for trials in English and Welsh general practice: implications for sample size calculations. *Stat.Med.* 2001; **20**(3):377-390.

8. Feng ZD, Thompson B. Some design issues in a community intervention trial. *Controlled Clinical Trials* 2002; **23**(4):431-449.

9. Klar N, Donner A. The merits of matching in community intervention trials: A cautionary tale. *Statistics in Medicine* 1997; **16**(15):1753-1764.

10. Donner A, Taljaard M, Klar N. The merits of breaking the matches: A cautionary tale. *Statistics in Medicine* 2007; **26**(9):2036-2051.

11. Evans BA, Feng Z, Peterson AV. A comparison of generalized linear mixed model procedures with estimating equations for variance and covariance parameter estimation in longitudinal studies and group randomized trials. *Stat.Med.* 2001; **20**(22):3353-3373.

12. Bennett S, Parpia T, Hayes R, Cousens S. Methods for the analysis of incidence rates in cluster randomized trials. *International Journal of Epidemiology* 2002; **31**(4):839-846.

13. Spiegelhalter DJ. Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* 2001; **20**(3):435-452.

14. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* 2001; **20**(3):453-472.

15. Ukoumunne OC. A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. *Stat.Med.* 2002; **21**(24):3757-3774.

16. Heo M, Leon AC. Comparison of statistical methods for analysis of clustered binary observations. *Statistics in Medicine* 2005; **24**(6):911-923.

17. Heo M, Leon AC. Performance of a mixed effects logistic regression model for binary outcomes with unequal cluster size. *Journal of Biopharmaceutical Statistics* 2005; **15**(3):513-526.

18. Braun TM. A mixed model-based variance estimator for marginal model analyses of cluster randomized trials. *Biometrical Journal* 2007; **49**(3):394-405.

19. Ukoumunne OC, Carlin JB, Gulliford MC. A simulation study of odds ratio estimation for binary outcomes from cluster randomized trials. *Statistics in Medicine* 2007; **26**(18):3415-3428.

20. Taljaard M, Donner A, Klar N. Imputation strategies for missing continuous outcomes in cluster randomized trials. *Biometrical Journal* 2008; **50**(3):329-345.

21. Ukoumunne OC, Forbes AB, Carlin JB, Gulliford MC. Comparison of the risk difference, risk ratio and odds ratio scales for quantifying the unadjusted intervention effect in cluster randomized trials. *Statistics in Medicine* 2008; **27**(25):5143-5155.

22. Jo B, Asparouhov T, Muthen BO. Intention-to-treat analysis in cluster randomized trials with noncompliance. *Statistics in Medicine* 2008; **27**(27):5565-5577.

23. Campbell MK, Elbourne DR, Altman DG. The CONSORT statement for cluster randomised trials. *Medicina Clinica* 2005; **125**:28-31.

24. Donner A, Klar N. Methods for Comparing Event Rates in Intervention Studies When the Unit of Allocation Is A Cluster. *American Journal of Epidemiology* 1994; **140**(3):279-289.

25. Young M.L. *Generalized estimating equations (GEE) with design-based correlation structures for cluster-unit trials*. University of North Carolina: 2003.

26. Zeger SL, Liang KY, Albert PS. Models for Longitudinal Data - A Generalized Estimating Equation Approach. *Biometrics* 1988; **44**(4):1049-1060.

27. Horton NJ, Lipsitz SR. Review of software to fit generalized estimating equation regression models. *American Statistician* 1999; **53**(2):160-169.

28. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? *Stat.Med.* 2009; **28**(2):221-239.

29. Young ML, Preisser JS, Qaqish BF, Wolfson M. Comparison of subject-specific and population averaged models for count data from cluster-unit intervention trials. *Statistical Methods in Medical Research* 2007; **16**(2):167-184.

30. Litiere S, Alonso A, Molenberghs G. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine* 2008; **27**(16):3125-3144.

31. Congdon P. *Bayesian Statistical Modelling*. John Wiley & Sons: Chichester, 2006.

32. Bernardo JM, Villegas MAG, Lindley DV, Schervish MJ. Objective Bayesian point and region estimation in location-scale models. *Sort-Statistics and Operations Research Transactions* 2007; **31**(1):3-+.

33. Hurton P.R. *Waterborne disease epidemiology and ecology*. John Wiley & Sons: 1997.

34. WHO/UNICEF. Join Monitoring Programme for Water Supply and Sanitation. Global water supply and sanitation assessment. 2000. Geneva, World Health Organization.
Ref Type: Report

35. Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *BMJ* 2007.

36. Dejung S, Fuentes I, Almanza G, Jarro R, Navarro L, Arias G, Urquieta E, Torrico A, Fenandez W, Iriarte M, Birrer C, Stahel WA, Wegelin M. Effect of solar water disinfection (SODIS) on model microorganisms under improved and field SODIS conditions. *Journal of Water Supply Research and Technology-Aqua* 2007; **56**(4):245-256.

37. Morris SS, Cousens SN, Lanata CF, Kirkwood BR. Diarrhoea--defining the episode. *Int J Epidemiol* 1994; **23**(3):617-623.

38. Wright JA, Gundry SW, Conroy R, Wood D, Du PM, Ferro-Luzzi A, Genthe B, Kirimi M, Moyo S, Mutisi C, Ndamba J, Potgieter N. Defining episodes of diarrhoea: results from a three-country study in Sub-Saharan Africa. *J Health Popul Nutr* 2006; **24**(1):8-16.

39. Brown H, Prescott R. *Applied Mixed Models in Medicine*. Wiley & Sons: Ontario, 2001.

40. SAS Institute Inc. *SAS/STAT 9.1 user's guide*. SAS institute Inc.: Cary: NC, 2004.

41. SAS Institute Inc. *The GLIMMIX Procedure*. SAS Institute Inc.: Cary, North Carolina, USA, 2006.

42. Chin HC, Quddus MA. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis and Prevention* 2003; **35**(2):253-259.

43. Tseloni A. Multilevel modelling of the number of property crimes: household and area effects. *Journal of the Royal Statistical Society Series A-Statistics in Society* 2006; **169**:205-233.

Chapter II:

A cluster-randomized, controlled trial of solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia

Daniel Mäusezahl[1], Andri Christen[1], Gonzalo Durán Pacheco[1], Fidel Alvarez Tellez[2], Mercedes Iriarte[3], Maria E. Zapata[3], Myriam Cevallos[1], Jan Hattendorf[1], Monica Daigl Cataneo[1], Benjamin Arnold[4], Thomas Smith[1], John M. Colford, Jr[4].

[1] *Department of Public Health and Epidemiology, Swiss Tropical Institute, University of Basel, Switzerland*
[2] *Project International Concern (PCI) Cochabamba, Bolivia*
[3] *Universidad Mayor de San Simón,Cochabamba, Bolivia*
[4] *Division of Epidemiology, School of Public Health, University of California, Berkeley*

This paper has been published in PLoS Med 6(8): e1000125.
doi:10.1371/journal.pmed.1000125

Abstract

Background: Solar drinking water disinfection (SODIS) is a low-cost, point-of-use water purification method that has been disseminated globally. Laboratory studies suggest that SODIS is highly efficacious in inactivating waterborne pathogens. Previous field studies provided limited evidence for its effectiveness in reducing diarrhoea.

Methods and findings: We conducted a cluster-randomized controlled trial in 22 rural communities in Bolivia to evaluate the effect of SODIS in reducing diarrhoea among children under the age of 5 y. A local nongovernmental organisation conducted a standardised interactive SODIS-promotion campaign in 11 communities targeting households, communities, and primary schools. Mothers completed a daily child health diary for 1 y. Within the intervention arm 225 households (376 children) were trained to expose water-filled polyethyleneteraphtalate bottles to sunlight. Eleven communities (200 households, 349 children) served as a control. We recorded 166,971 person-days of observation during the trial representing 79.9% and 78.9% of the total possible person-days of child observation in intervention and control arms, respectively. Mean compliance with SODIS was 32.1%. The reported incidence rate of gastrointestinal illness in children in the intervention arm was 3.6 compared to 4.3 episodes/year at risk in the control arm. The relative rate of diarrhoea adjusted for intracluster correlation was 0.81 (95% confidence interval 0.59–1.12). The median length of diarrhoea was 3 d in both groups.

Conclusions: Despite an extensive SODIS promotion campaign we found only moderate compliance with the intervention and no strong evidence for a substantive reduction in diarrhoea among children. These results suggest that there is a need for better evidence of how the well-established laboratory efficacy of this home-based water treatment method translates into field effectiveness under various cultural settings and intervention intensities. Further global promotion of SODIS for general use should be undertaken with care until such evidence is available.

Introduction

Globally, 1.8 million people die every year from diarrhoeal diseases the vast majority of whom are children under the age of 5 y living in developing countries [1]. Unsafe water, sanitation, and hygiene are considered to be the most important global risk factors for diarrhoeal illnesses [2].

Recent systematic reviews concluded that interventions to improve the microbial quality of drinking water in households are effective at reducing diarrhoea, which is a principal source of morbidity and mortality among young children in developing countries [3–5]. One widely promoted water disinfection method with encouraging evidence of efficacy in laboratory settings is solar drinking water disinfection (SODIS) [6]. Global efforts are underway to promote SODIS as a simple, environmentally sustainable, lowcost solution for household drinking water treatment and safe storage (www.who.int/household_water, www.sodisafricanet.org). SODIS is currently promoted in more than 30 countries worldwide (www.sodis.ch) and in at least seven Latin American countries through the SODIS Foundation including in Bolivia.

Despite this widespread promotion, evidence of the effectiveness of SODIS from field studies is limited. The three reported SODIS trials to date implemented the intervention at the household level, two of them in highly controlled settings that ensured very high compliance [7–9]. The highest reduction in incidence (36%) was recorded in a trial carried out among 200 children in an urban slum in Vellore, India [9].

Because SODIS is a behavioural intervention designed to reduce infectious diarrhoea, disease transmission and its interruption likely have community level dynamics [10]. In addition, because SODIS is typically rolled out in practice through community rather than household level promotion, there is an urgent need for effectiveness data from such settings. We conducted a community-randomized intervention trial to evaluate the effectiveness of SODIS in decreasing

diarrhoea in children < 5 y in rural communities in Bolivia.

Methods

Ethics Statement

The study was approved by the three human subjects review boards of the University of Basel, Switzerland, the University of California, Berkeley, and the University of San Simon, Cochabamba, Bolivia. The Cochabamba and Totora municipal authorities also approved the study and informed consent was obtained from community leaders and male and female household heads prior to implementation of the study. Informed consent was obtained before randomisation to the treatment arms (Figure II.1). Mildly ill children from households participating in the study were provided with and instructed to use oral rehydration salts, or they were referred by field staff to the local health system where clinical services were provided free of charge. The project provided transport and treatment costs for those patients. All project staff completed training on research ethics (www.fhi.org/training/sp/Retc/). Project staff comprised all project personnel of all project partners. Field staff comprised all personnel working in our laboratories and at our Totora field station including data enumerators and data- and project-management staff, supervisors, and community-based field workers living in the study communities. The trial protocol (Text S1) and the CONSORT statement checklist (Text S2) are available online as supporting information.

Site and Population

Our trial, the Bolivia Water Evaluation Trial (BoliviaWET), was conducted in an ethnically homogeneous Quechua setting in rural Totora District, Cochabamba Department, Bolivia. Our study was part of a comprehensive SODIS roll-out programme in collaboration with Project Concern International, a nongovernmental organisation (NGO). Most of the local residents are farmers, typically living in small compounds of three buildings with mud floors, with five or more persons sleeping in the same room. Our own surveys showed that 15% of homes have a latrine or other sanitary facilities and that most residents defecate in the nearby environment.

Drinking water is typically stored in 10-l plastic buckets or open jerry cans of 5–20 l in the household. Baseline assessments of the drinking water quality in the home indicated a median contamination of thermotolerant coliforms (TTC) of 32 TTC/

100 ml (interquartile range (IQR)= 3–344; n = 223). Samples of at least one water source per community were tested for Giardia lamblia and Cryptosporidium parvum. The two parasites were detected in 18/24 and 11/23 water samples, respectively.
Parasites were detected by using immunomagnetic separation and PCR techniques [11]. Piped water, when available, is not chlorinated.

Design

Twenty-seven of 78 communities in the study area fulfilled the selection criteria (geographically accessible all year round; at least 30 children < 5 y; reliance on contaminated drinking water sources). Two communities were excluded because of other ongoing health and hygiene campaigns, and three communities withdrew participation before baseline activities because of a change in political leadership. Community health workers undertook a census and identified households with at least one child < 5 y. All children < 5 y were enrolled in the participating villages.

We pair-matched communities on the incidence of child diarrhoea as measured in an 8-wk baseline survey [12]. The intervention was then assigned randomly to one community within each of the 11 consecutive pairs. This assignment was done during a public event because key political stakeholders were worried about possible backlash, public outcry, or a drop-off in group participation, which would result from providing some members with a new benefit while others got ‘‘nothing.’’ It was agreed that a public drawing event was necessary to increase perceived fairness among the participating district and municipal authorities. Three authorities, the district head (Alcalde), representatives of the Ministries of Health and Education, and the deputy of the farmers union (Central Campesina), each drew one of two balls (with community codes inscribed that were randomly assigned beforehand) representing paired communities from a concealed box. It was agreed that the first draw assigned the community to the intervention arm. The group allocation was immediately recorded in a protocol by an independent witness. Subsequently, the witness disclosed the sequence, informed the community members and the authorities present in the town hall, and all drawers signed the protocol.

We explicitly chose community-level randomization because important components of the intervention (i.e., community efforts to encourage adoption of the SODIS-method) would occur at the community level. Randomization below the community level would not reflect the reality of scale-up programme implementation, and we would not have captured the potential community-level reinforcement of the behaviour change. Furthermore, community-level randomization is considered ethically optimal, because participants expect to equally benefit from interventions within their community [13–15]. Additionally, we believed cross-contamination (of the intervention) between the intervention and control communities was minimised by vast geographical dispersion of the communities. Control communities knew from the beginning of the study that they would receive the intervention as part of the NGO's development plans after study completion. It was not possible for the NGO to carry out the intervention in all the communities at the same time, thus making randomization feasible and acceptable to the three ethical review boards overseeing the study.

Sample size was calculated according to methods outlined by Hayes and Bennett [16], assuming an incidence rate (IR) in the control villages of five episodes/child/year [17], and accounting for clustering, the number of episodes, and the expected effect. We assumed a coefficient of between-cluster variation (k) of similar studies, between 0.1–0.25 (as cited by Hayes and Bennett) and a minimum of 10 child-years of observation per cluster [16]. We calculated that nine pairs of clusters were required to detect a difference of at least 33% in the IR between the control and intervention arms with 80% power, k= 0.20 and an alpha level of 0.05. Anticipating a drop-out of at least one cluster per arm and a loss of follow-up of individuals, the final sample size was adjusted to 11 pairs with 30 children per community cluster. We powered the study to detect a 33% reduction in diarrhoea incidence after reviewing the evidence base for point-of-use water treatment at the time of the study's inception in 2002 [18].

Implementation of the intervention

The SODIS intervention was designed according to the published guidelines for national SODIS dissemination (http://www.sodis.ch/files/TrainingManual_sm.pdf). Promotion activities were targeted at primary caregivers and all household members (biweekly), whole communities (monthly), and primary schools (three times) by the

NGO as part of its regional community development programme. Eleven communities (262 households and 441 children) were randomized to the intervention; 11 communities (222 households, 378 children) served as a control group (Figure II.1). The implementation scheme and detailed description of the intervention in the intervention arm (and the control arms after study end) are described in Appendix B. For a period of 15 mo an intensive, standardised, and repeated interactive promotion of the SODIS method was implemented in the intervention communities beginning 3 mo before the start of follow-up.

Within the intervention arm, participating households were supplied regularly with clean, recycled polyethyleneteraphtalate (PET) bottles. The households were taught through demonstrations, role plays, video, and other approaches to expose the water-filled bottles for at least 6 h to the sun. NGO staff emphasized the importance and benefits of drinking only treated water (especially for children), explained the germ–disease concept, and promoted hygiene measures such as safe drinking water storage and hand washing as they relate to the understanding of drinking water and the faecal–oral route of transmission of pathogens (Appendix B). During household visits the NGO staff encouraged all household members to apply the method, answered questions, and assisted mothers and primary caregivers to integrate the water treatment into daily life. The same intervention (in terms of contents and messages) was supplied to the communities in the control arm by the NGO-staff at the end of the study (Appendix B).

Outcome

The primary outcome was the IR of diarrhoea among children <5 y, defined as number of diarrhoea episodes per child per year obtained from daily assessment of individual diarrhoea occurrence. We applied the WHO definition for diarrhoea of three or more watery bowel movements or at least one mucoid/bloody stool within 24 h [19,20]. We defined a new episode of diarrhoea as the occurrence of diarrhoea after a period of 3 d symptom-free [20–22]. An episode of diarrhoea was labelled ‘‘dysentery’’ if signs of blood or mucus in the stool were recorded at any time. We also calculated the longitudinal prevalence (number of days a child suffered diarrhoea divided by the number of days of observation) because of its closer relation to severity,

growth faltering, and mortality than diarrhoea incidence [19,23]. Severe diarrhoea was defined as the occurrence of diarrhoea on more than 10% of the observed days [24].

Data collection and field staff

The primary outcome was measured by community-based field workers who were recruited nearby and who lived one per community during data collection periods. The field workers were extensively trained in interviewing and epidemiological observation techniques, data checking, recording, and in general approaches to community motivation. Community-based field workers were randomly rotated between communities every 3 mo. Child morbidity was reported by the closest caregiver using the vernacular term ‘‘K’echalera,’’ which had been established previously to correspond to the WHO definition of diarrhoea [25]. Mothers or closest caretakers kept a 7-d morbidity diary recording daily any occurrence of diarrhoea, fever, cough, and eye irritations in study participants [25]. Community-based field workers visited households weekly to collect the health diaries, and supervisors revisited an average 7% of homes. Discrepancies between supervisors and community-based field workers' records were clarified during a joint home revisit. Child exposure risks were also assessed by community-based staff interviewing mothers once during baseline and twice during the 1-y follow-up.

Compliance with the SODIS method was measured using four different subjective and objective indicators. Three of the indicators were assessed by field staff independent from the implementing NGO: (i) the number of SODIS-bottles exposed to sunlight and, (ii) the number of bottles ready-to-drink in the living space, and (iii) the personal judgment about families' user-status was provided by community-based field workers living among the families in the intervention arm. Judgement criteria for this main compliance indicator study included observing regular SODIS practice and bottles exposed to sun or ready to drink in the kitchen and being offered SODIS-treated water upon request. The fourth SODIS-use indicator was based on self-reporting and caregivers' knowledge of and attitudes toward the intervention that was assessed at the beginning (i.e., 3 mo after start of the intervention) and at the end of the 12-mo follow-up period.

Statistical Analysis

An intention-to-treat analysis was applied comparing the IR of diarrhoea between children ,5 y in intervention and control communities. Diarrhoea prevalence (PR) and severe diarrhoea (SD) were additionally analysed. Generalized linear mixed models (GLMM) were fitted to allow for the hierarchical structure of the study design (pair-matched clusters). In contrast to our original trial protocol we selected the GLMM approach rather than generalized estimating equations (GEE) because recent publications indicated that the latter method requires a larger number of clusters to produce consistent estimates [26].

The crude (unadjusted) model included only the design factors and the intervention effect [12,27]. Further models included potential confounders (selected a priori: child's age, sex, child hand-washing behaviour, and water treatment at baseline). Following an evaluation of the best fit, the GLMM included the log link function for negative binomial data (IR) and logit for binomial data (PR and SD). Denoting the link function of the outcome Y by g(E(Y)), the crude and adjusted models were: $g(E(Y_{ijk})) = \mu + B_i + \tau_j + \xi_{ij}$, and $g(E(Y_{ijk})) = \mu + Bi + \tau_j + \xi_{ij} + \mathbf{x'b}$ where Y_{ijk} denotes the observed outcome value for the k-th individual from a community allocated to the j-th intervention, in the i-th pair, μ is the general mean, B_i is the random effect of the i-th pair ~ $N(0, \sigma_p^2)$, τ_j is the fixed effect of the SODIS intervention, and ξ_{ij} is the random effect of the interaction of the i-th pair with the j-th intervention applied to the community ~ $N(0, \sigma_{pc}^2)$ (signifying the within-pair cluster variance and used as error term for τ_j), **x** is the vector of potential confounding factors and **b** the vector of the corresponding regression coefficients.

The intracluster correlation coefficient (ICC) and the coefficient of between-cluster variation (k) were calculated after data collection to validate the degree of clustering and our assumptions for the sample size. ICC and k were estimated from the unscaled variance of the IR's GLMM. To estimate the uncertainty of ICC and k, we obtained the 95% credible region (Bayesian equivalent of 95% confidence interval [CI]) through an analogous Bayesian hierarchical regression [28]. Noninformative priors were used. The statistical analyses were performed using SAS software v9.1 (PROC GLIMMIX, SAS Institute Inc.) and WinBUGS v1.4 (Imperial College and MRC).

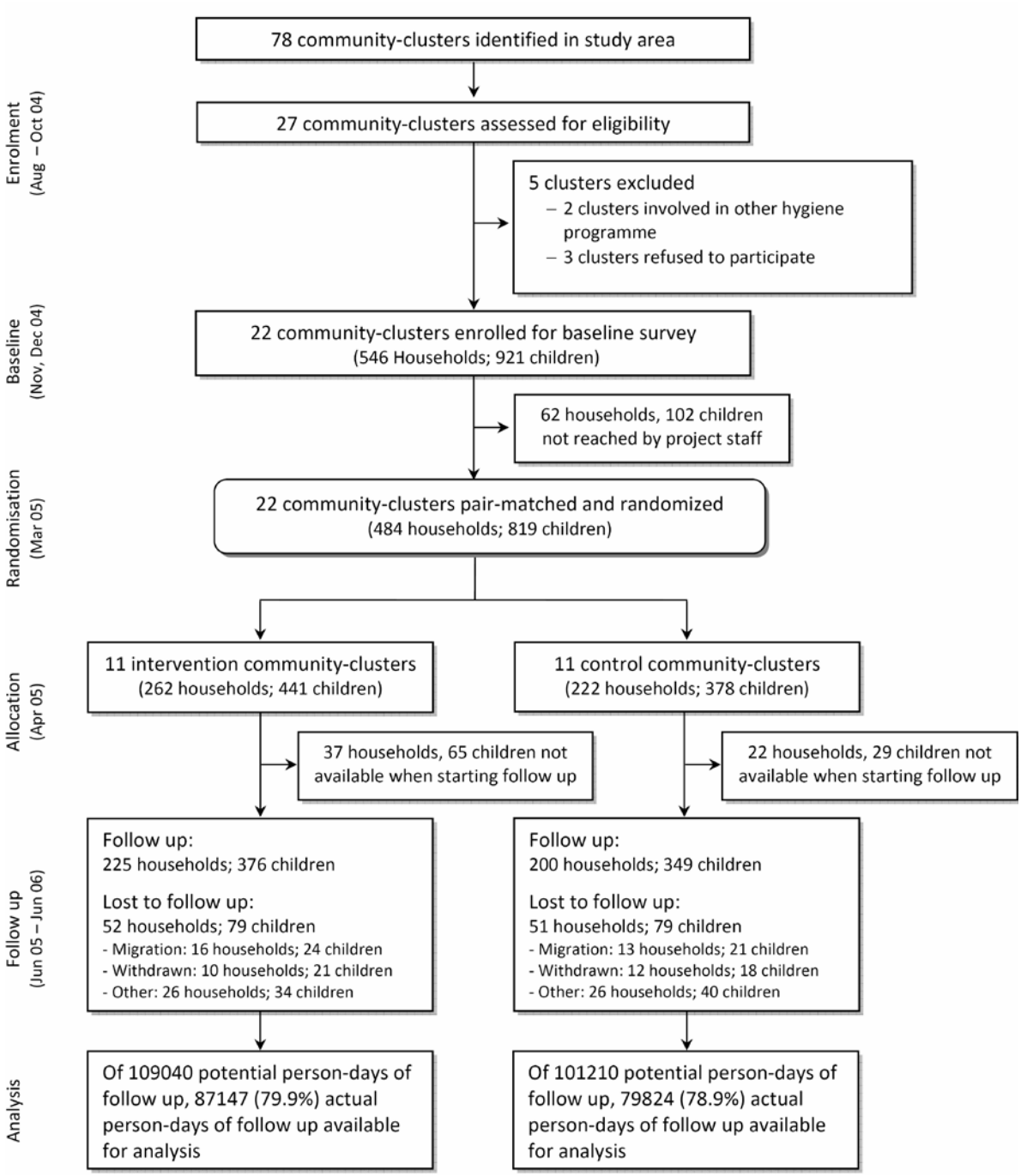


Figure II.1: Community-randomized trial flow diagram on point-of-use solar water disinfection in totora district, bolivia.

Results

Participant flow and recruitment

Among the 1,187 households in the 22 communities there were 546 that met the inclusion criteria (Figure II.1). The median number of participating households with

children <5 y per community was 22. Because of political unrest and national election campaigns in 2005 a period of 6 mo passed between the baseline and the start of follow-up. Subsequently, 62 households (102 children) were no longer traceable before randomisation, and 59 households (37 intervention, 22 control) were lost before data collection had started. The loss to follow-up was balanced in intervention and control arms. Data were obtained from 376 children (225 households) in the intervention and 349 children (200 households) in the control arm, thus reaching our originally planned sample size.

Follow-up started in June 2005 and ended in June 2006. During the 51 wk of the study, information on the occurrence of diarrhoea was collected for 166,971 person-days representing 79.9% and 78.9% of the total possible person-days of child observation in intervention and control arms. We excluded from the potential observation time the experience of 94 children who dropped out before the start of follow-up. National festivities, holidays, and political unrest over the entire year amounted to further 9 wk during which outcome surveillance needed to be suspended. The main reasons for incomplete data collection were migration (28%) and withdrawal (67%). Supervisors reevaluated the outcome during 984 unannounced random home visits, and discrepancies between community-based field workers' and supervisors' records were found for five (0.5%) of all visits.

Baseline characteristics

At baseline the households in the different study arms were well balanced on multiple other factors suggesting successful randomisation (Table II.1). The main types of water sources for household chores and drinking were similar in both arms as was the distance to the source (median distance 50 m and 30 m in the control and intervention arms, respectively). Storing water for longer than 2 d was more common among the intervention (26.8%) than the control arm (13.9%). Nearly 30% of all households reported treating water regularly before drinking. Boiling was the most common water treatment before the trial (20.2% in both arms).

Table II. 1: Baseline community- and household characteristics of a community-randomized trial of SODIS

Characteristic		Control 11 clusters		Intervention 11 clusters
Demography				
Community size: Nr of households [Mean (sd)]		50 (20)		58(20)
Household size: Nr of household members [Mean (sd)]	N= 222	6.2 (2.1)	N= 262	6.3 (2.6)
Nr of children< 5 per household [Mean (sd)]		1.8 (0.7)		1.7 (0.8)
Nr of children< 5 per community [Mean (sd)]		35.3 (6.6)		41.4 (9.9)
Female household head [Nr (%)]		20 (9.0)		14 (5.4)
Closest child caregiver (female)		223 (99.5)		266 (99.6)
Age of closest child caregiver (yr) [Mean (sd)]		31(9)		30(10)
Nr of children <1		65 (4.7)		67 (4.1)
Nr of children <5		369 (26.6)		426 (25.9)
Education				
Household chief: Reported years of education [Mean (sd)]	N= 167	4.1 (2.6)	N= 178	4.2 (2.4)
Closest child caregiver: Reported years of Education [Mean sd)]	N= 179	2.5 (1.9)	N= 198	2.7 (1.8)
Socio-economic Variables				
Main occupation of the household chief as farmer	N= 208	180 (86.5)	N= 228	207 (90.8)
Ownership of truck, car or motorbike		12 (5.8)		14 (6.2)
Ownership of radio		129 (86.1)		194 (85.1)
Ownership of bicycle		109 (52.4)		121 (53.1)
Ownership of television		24 (11.5)		15 (6.6)
Nr of rooms in the house [Mean (sd)]		2.9 (1.4)		2.8 (1.2)
Water Management & Consumption				
Spring as source of drinking water	N= 208	100 (48.1)	N= 228	136 (59.6)
Tap as source of drinking water		108 (51.9)		129 (56.6)
River as source of drinking water		46 (22.1)		29 (12.7)
Rain as source of drinking water		31 (14.9)		71 (31.1)
Dug well as source of drinking water		31 (14.9)		37 (16.2)
Distance to water source (m) [Median (Q1, Q3)]		50 (7.5, 100)		30 (6, 150)
Container for water collection: Plastic bucket		189 (90.9)		205 (89.9)
Container for water collection: Jerry can		165 (79.3)		156 (68.4)
Container for water collection: Bottles		32 (15.4)		36 (15.8)
Container for water collection: Jar / Pitcher		13 (6.3)		20 (8.8)
Container for water collection: Barrel		10 (4.8)		25 (10.9)
Child's consumption of untreated water (glasses/day) [Mean (sd)]	M= 318	1.2 (1.2)	M= 359	1.2 (1.4)
Treat water before drinking	N= 208	59 (28.4)	N= 228	67 (29.4)
Store water for >2 days		29 (13.9)		61 (26.8)
Water storage container: Jerry can		23 (11.1)		49 (21.5)
Water storage container: Plastic bucket		17 (8.2)		37 (16.2)
Water turbidity in water storage container >30 NTU		13 (11.2)		24 (18.8)

Characteristic		Control 11 clusters		Intervention 11 clusters
Sanitation				
Reported Nr of interviewee's hand washing per day [Mean (sd)]	N= 177	3.8 (1.7)	N= 200	4.1 (1.8)
Reported Nr of child hand washing per day [Mean (sd)]	M= 348	2.5 (1.2)	M= 376	2.6 (1.4)
Child washes hands : Before eating		228 (65.5)		270 (71.8)
Child washes hands : When hands are dirty		62 (17.8)		56 (14.9)
Child washes hands : Other occasions		58 (16.7)		50 (13.3)
Latrine present	N= 208	27 (13.0)	N= 228	38 (16.7)
Use of latrine by the interviewee (day or night)		15 (7.2)		20 (8.8)
Feces visible in yard	N= 202	121 (59.9)	N= 219	124 (56.6)

Data shows numbers and percentages (%) unless otherwise specified
N = Number of households, M = Number of children
NTU: Nephelometric units, 30NTU: threshold for efficacious pathogen-inactivation of the SODIS method
Baseline data from Dec. 2004

Intervention and attendance

The NGO conducted 210 community events and 4,385 motivational household visits in intervention communities; 3,060 visits occurred in the households with children < 5 y followed up and analysed for the study, and 1,325 household visits took place in homes that were not taking part in the study. Study households attended a median of nine community events (IQR= 5–12) and were visited by the SODIS-programme team a median 11 times at home (IQR =7–18). To ensure a sufficient number of PET bottles, the NGO provided as many SODIS-bottles as required by participants (mean 955 bottles/community).

Compliance

Community-based field workers who were living in the communities throughout the study observed a mean SODIS-user rate of 32.1% in the intervention arm (minimum 13.5%, maximum 46.8%, based on their personal judgement) (Figure II.3). The mean proportion of households with SODIS-bottles exposed to the sun was 5 percentage points higher than the assessment by community-based field workers. In contrast, almost 80% of the households reported using SODIS at the beginning and end of the follow-up. About 14% of the households used the method more than two-thirds (> 66%) of the weeks during observation, and 43% of the households applied SODIS in more than 33% of the observed weeks (Table II.4).

Table II.2: Diarrhoea episodes, length of illness and days ill with diarrhoea

		N	Control	N	Intervention
Diarrhoea Illness Overview		**Children**		**Children**	
Days under observation	median (Q1, Q3)	349	263 (213, 274)	376	263 (222, 273)
Days at risk	median (Q1, Q3)	349	246 (192, 265)	376	247 (202, 265)
Nr of episodes	median (Q1, Q3)	349	1 (0, 3)	376	1 (0, 3)
Nr of dysentery episodes	median (Q1, Q3)	349	1 (0, 2)	376	1 (0, 2)
Days spent ill	median (Q1, Q3)	349	4 (0, 11)	376	4 (0, 12)
Episode length (days)	median (Q1, Q3)	349	3 (1, 5)	376	3 (2, 5)
Days under observation	Total		79'829		87'140
Days at risk	Total		75'077		82'682
Nr of episodes	Total		887		808
Nr of dysentery episodes	Total		460		431
Days spent ill	Total		3111		3038
Diarrhoea Incidence	**Age class**	**Children**	**Inc. Rate**	**Children**	**Inc. Rate**
Nr episodes / (child x year at risk)	<1	16	7.8	15	11.1
	1 - 2	67	7.1	70	5.5
	2 - 3	67	4.3	82	3.8
	3 - 4	77	3.2	75	2.8
	4 - 5	71	3.4	80	2.1
	5 - 6	50	2.7	53	2.5
	Total*	**349**	**4.3**	**376**	**3.6**
Diarrhoea Prevalence	**Age class**	**Children**	**Mean (std)**	**Children**	**Mean (std)**
Nr days ill / (child x year)	<1	16	27.4 (28.3)	15	42.3 (40.7)
	1 - 2	67	31.4 (42.2)	70	23.0 (26.1)
	2 - 3	67	19.0 (47.5)	82	16.4 (28.4)
	3 - 4	77	11.7 (24.5)	75	7.3 (9.7)
	4 - 5	71	9.5 (15.1)	80	6.2 (12.4)
	5 - 6	50	6.9 (11.8)	53	7.7 (10.4)
	Total*	**349**	**16.5 (32.8)**	**376**	**13.5 (22.4)**
Diarrhoea Illness	**Days spent ill**	**Children**	**%**	**Children**	**%**
	0 days	97	27.8	126	33.5
	1 - 2 days	50	14.3	42	11.2
	3 - 7 days	91	26.1	80	21.3
	8 - 14 days	49	14.0	59	15.7
	15 - 21 days	27	7.7	33	8.8
	22 - 40 days	18	5.2	21	5.6
	> 40 days	17	4.9	15	4.0
	Total	**349**	**100**	**376**	**100**
Diarrhoea Illness Duration	**Episode duration**	**Episodes**	**%**	**Episodes**	**%**
	1 day	250	28.2	191	23.6
	2 - 3 days	303	34.2	292	36.1
	4 - 7 days	258	29.1	250	30.9
	8 - 13 days	54	6.1	59	7.3
	14+ days	22	2.5	16	1.9
	Total	**887**	**100**	**808**	**100**
Prevalence of Other Symptoms					
[days / (child x year)]		**Children**	**Mean (std)**	**Children**	**Mean (std)**
Vomit		349	5.5 (13.2)	376	4.0 (8.9)
Fever		349	21.0 (33.0)	376	15.1 (19.8)
Cough		349	41.9 (48.3)	376	30.9 (39.4)
Eyes irritation		349	12.8 (29.8)	376	8.3 (19.5)

* includes one child per treatment arm with unknown age

Diarrhoeal illness in the control and intervention arm

No positive effect of compliance (proportion of weeks of observed SODIS use) on the IRs in the intervention arm was observed. The incidence did not decline with the increase of weeks using SODIS (Figure II.4). Seasonal variation in compliance was observed. The proportion of SODIS-practising households was consistently below average during weeks 4–16 (January 2005–April 2006), which corresponded to the labour intensive cultivating period from November to May.

The median proportion of sunny days with more than 6 h of sunshine was 70.2% and 67.2% in intervention and control communities, respectively, consistent with the technical and climatic conditions necessary for the proper functioning of the ultraviolet SODIS purification process [29] during the study (Table II.4).

A multivariable model adjusting for age, sex, baseline-existing water treatment practises and child hand-washing was consistent in its estimate of effect. (RR=0.74, 95% CI 0.50-1.11). We repeated the analysis by including confounding covariates in the order of occurrence of the variables in Table II.3 to confirm that the conclusions were not sensitive to the choice of covariates. None of the models yielded significant results for the effect of SODIS (all p-values >0.1) or resulted in meaningful changes in estimates of relative rates or odd ratios. Figure II.2 shows the relationship between study time and diarrhoea in the control and intervention arm. We found no statistically significant effect of the interaction of time and intervention in a time-dependent model.

Table II.3: Effect of SODIS on diarrhoea episodes, longitudinal prevalence, severe diarrhoea, and dysentery episodes.

Outcome	Model	Nr of Children	Parameter	Relative Rate/Odds ratio	95% CI	p value
Nr of Episodes (RR)	Unadjusted	725	Intervention	0.81	(0.59, 1.12)	.19
	Adjusted	644	Intervention	0.74	(0.50, 1.11)	.14
			age	0.75	(0.70, 0.81)	<.001
			sex	1.03	(0.84, 1.26)	.80
			water treatment	1.05	(0.81, 1.36)	.69
			hand washing	0.93	(0.85, 1.02)	.13
Prevalence (OR)	Unadjusted	725	Intervention	0.92	(0.66, 1.29)	.62
	Adjusted	644	Intervention	0.91	(0.64, 1.30)	.60
			age	0.67	(0.61, 0.73)	<.001
			sex	1.05	(0.84, 1.31)	.68
			water treatment	1.00	(0.76, 1.33)	.97
			hand washing	0.94	(0.84, 1.04)	.23
Severe Diarrhoea (OR)	Unadjusted	643	Intervention	0.91	(0.51, 1.63)	.75
	Adjusted	589	Intervention	1.02	(0.52, 2.01)	.95
			age	0.52	(0.40, 0.67)	<.001
			sex	1.12	(0.63, 2.01)	.69
			water treatment	1.59	(0.81, 3.12)	.18
			hand washing	0.94	(0.75, 1.19)	.62
Dysentery (OR)	Unadjusted	725	Intervention	0.80	(0.55, 1.17)	.23
	Adjusted	644	Intervention	0.75	(0.47, 1.18)	.20
			age	0.73	(0.67, 0.80)	<.001
			sex	1.00	(0.80, 1.26)	.97
			water treatment	1.15	(0.87, 1.53)	.33
			hand washing	0.91	(0.82, 1.01)	.06

Nr of episodes: Nr of episodes per days at risk
Prevalence: Nr of days ill per days under observation
Severe diarrhoea: Diarrhoea during >10% of all days (only children with more than 100 days of observation are included)
Unadjusted: General linear mixed models; only design factors and treatment are included
Adjusted: Effects of treatment and covariates
Sex: 0 = female, 1 = male; Water treatment: Water treatment at baseline, 0 = no treatment, 1 = treatment (Chlorination or Boiling or SODIS); Hand washing: Reported number of child's hand washing per day at baseline

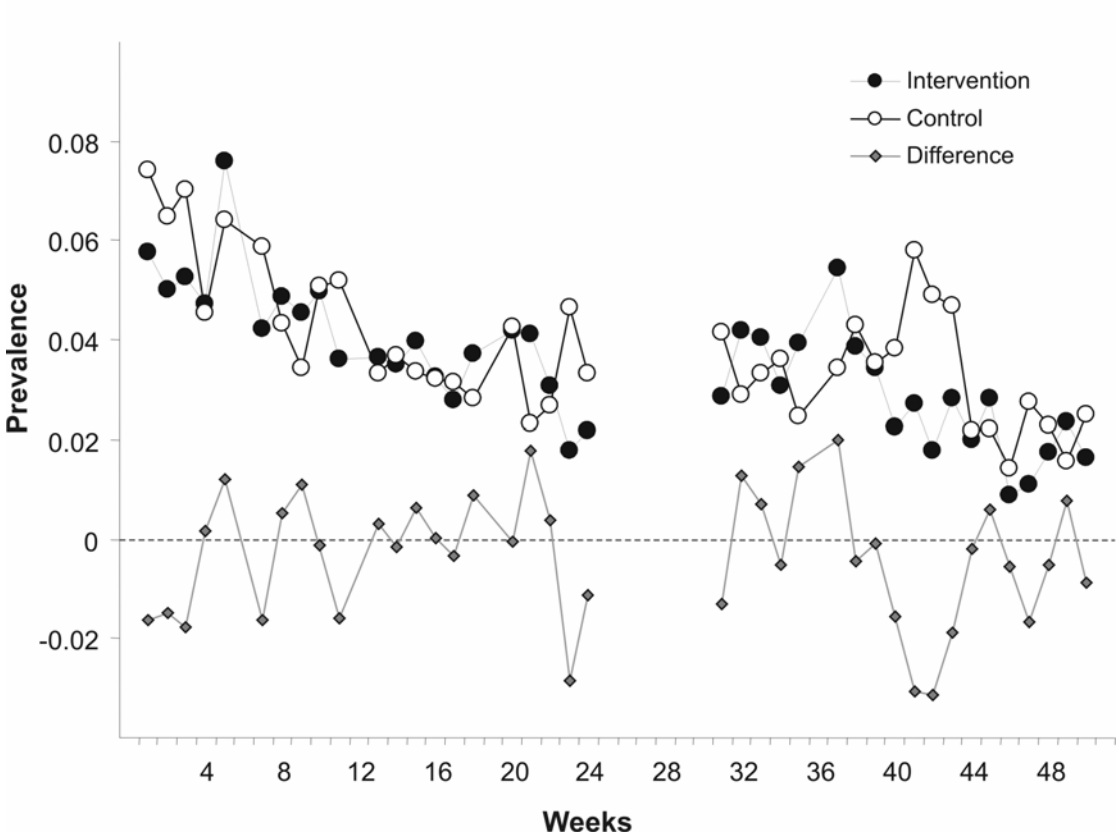
Diarrhoeal illness by compliance

No positive effect of compliance (proportion of weeks of observed SODIS use) on the IRs in the intervention arm was observed. The incidence did not decline with the increase of weeks using SODIS (Figure II.4). Seasonal variation in compliance was observed. The proportion of SODIS-practising households was consistently below

average during weeks 4–16 (January 2005–April 2006), which corresponded to the labour intensive cultivating period from November to May.

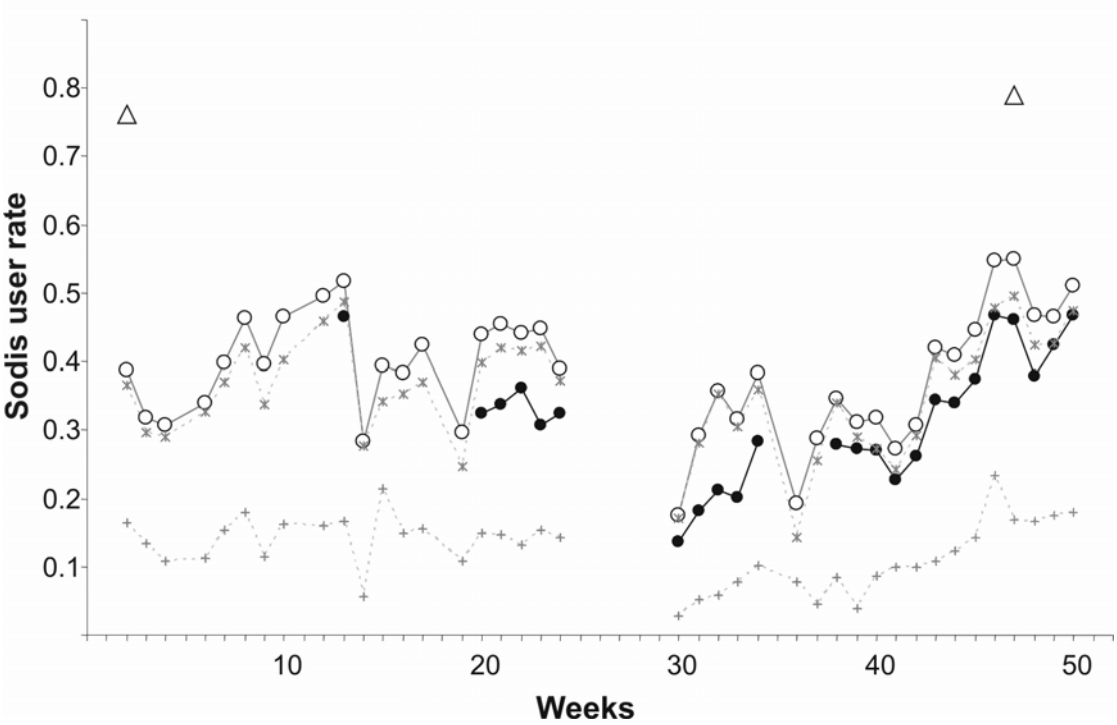
The median proportion of sunny days with more than 6 h of sunshine was 70.2% and 67.2% in intervention and control communities, respectively, consistent with the technical and climatic conditions necessary for the proper functioning of the ultraviolet SODIS purification process [29] during the study (Table II.4).

Figure II.2: weekly prevalence of child diarrhoeal illness.



Legend: Weekly points are derived from daily prevalence data of each participating child

Figure II.3: Weekly observed proportion of households using solar water disinfection as point-of-use drinking water purification method.



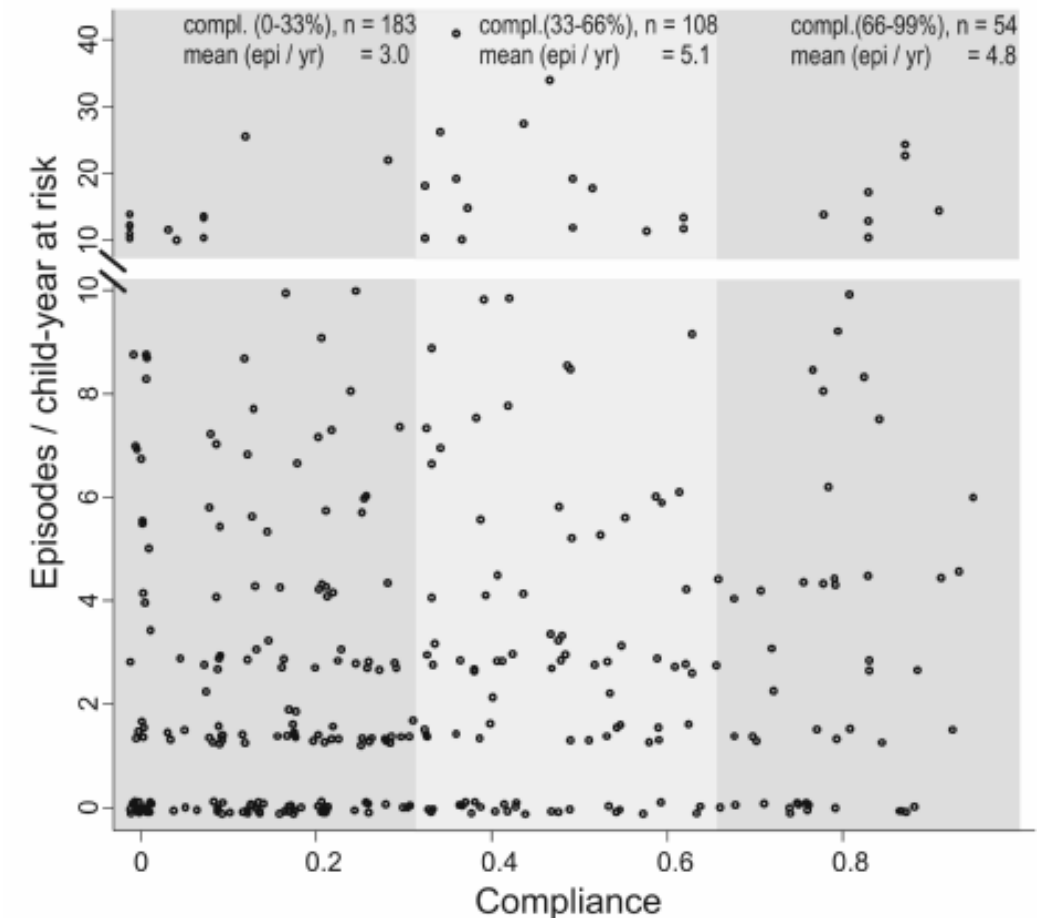
Legend: Open triangles: self-reported SODIS-use at the beginning (after 3 month of initial SODIS promotion) and at the end of follow-up; filled dots: SODIS-use observed by project staff living in the community (see methods for definition); open circles: SODIS bottles observed on the roof and/or in the kitchen; Stars: SODIS-bottles on the roof; crosses: SODIS-bottles in the kitchen.

Table II.4: Climatic Conditions and SODIS-use of a Cluster-randomized Trial Involving 22 Rural Communities of Totora District, Bolivia.

Description		**Control** (N= 11 clusters)	**Intervention** (N= 11 clusters)
Climate	Percentage of sunny days (>6hrs sunshine) [median of clusters (min, max)]	70 (57, 78)	67 (44, 77)
	Average duration of sunshine [median of clusters (min, max)]	7.0 (6.3, 8.0)	7.1 (4.5, 8.3)
SODIS-use	Observed level of SODIS use [a]	Percentage of households	Percentage of households
	0.66 - 1	0 %	14 %
	0.33 - 0.66	0.5 %	29 %
	0 - 0.33	99.5 %	57 %

[a] Proportion of weeks in which SODIS was used, as estimated by community-based project staff at the end of study. Households with less than 10 weeks of observation are excluded

Figure II.4: Compliance of Using Solar Water Disinfection (SODIS) and Child Diarrhoea in Rural Bolivia.



Legend: Compliance of SODIS use is estimated as the proportion of weeks a family has been classified as a SODIS user by community-based project staff. Dots: number of episodes per child-year at risk;. Small random noise was added to the dots to avoid over plotting. Only children with at least 110 days under observation are included.

Discussion

We conducted a community-randomized trial within the operations of an ongoing national SODIS-dissemination programme which provided an intensive training and repeated reinforcement of the SODIS-intervention throughout the study period. In this context of a 'natural experiment' we found a relative rate of 0.81 for the incidence rate of diarrhoea episodes among children assigned to SODIS compared to controls. However, the confidence interval included unity (RR=0.81, 95% CI 0.59–1.12) and therefore we conclude that there is no strong evidence for a substantive reduction. Subsequently, we discuss the primary outcome in context of other study findings, and

explain why we hypothesize that the true effect – if there is any – might be smaller. First, the estimate for the longitudinal prevalence of diarrhoea was substantially smaller (OR=0.92, 95% CI 0.66-1.29) then the estimate of incidence and there is some evidence that prevalence is a better predictor in terms of mortality and weight gain than incidence [23].

The absence of a time-intervention interaction in our time-dependent analysis suggested no increased health benefits with the ongoing intervention. Furthermore, within the intervention arm, there was no evidence that increased compliance was associated with a lower incidence of diarrhoea (Figure II.4). However, we interpret this post hoc subgroup analysis cautiously because compliant SODIS users might differ in important ways from noncompliant users. A compliant SODIS user might be more accurately keeping morbidity diaries, whereas less compliant families may tend to underreport diarrhoeal illness. Or, households with a high burden of morbidity might be more likely to be compliant with the intervention. Both of these scenarios could lead to an underestimation of the effectiveness of SODIS.

Further, analysing the laboratory results from 197 randomly selected stool specimens the proportion of *Cryptosporidium parvum* was lower in the intervention children (5/94 vs. 2/103), and other pathogens were found at similar proportions in intervention and control children (*Gardia lamblia*: 39/94 vs. 40/103; Salmonella sp.: 2/94 vs. 3/104; Shigella sp.: 3/94 vs. 3/104). In further exploring the occurrence of other illness symptoms we found the prevalence of eye irritations and cough to be lower in the intervention group compared to the control group. This difference could be the result of the limited hygiene component in the intervention that increased hygiene awareness among the treatment communities. An alternative explanation is that the lack of blinding led to biased (increased) health outcome reporting in the intervention group.

Due to the nature of the intervention neither participants nor personnel were blinded to treatment assignment. Ideally, blinding to the intervention allocation should apply to the NGO staff administering the SODIS intervention and our enumerators assessing

outcomes [30]. Although the former could not be blinded in our study (for obvious reasons), the latter would inevitably be able to identify the intervention status of the cluster through the visible display of bottles to sunlight in the village or directly at the study home during home visits. These problems are consistent with nearly all household water treatment interventions [5] and other public health cluster randomized trials [31,32]. Schmidt and Cairncross [33] recently argued that reporting bias may have been the dominant problem in unblinded studies included in a meta-analysis reporting a pooled estimate of a 49% reduction of diarrhoea in trials investigating the effects of drinking water quality interventions [5]. However, their review of only four available blinded trials showing no effect demonstrates weak support for contrast. In addition, all of the blinded trials exhibited analytical shortcomings or had very broad CIs suggesting very low power. In the absence of blinding—unavoidable inmany behavioural change interventions or household water treatment studies—we believe that data collection independent from the implementation is a crucial factor. Future reviews should include reporting on such additional quality parameters.

In our study the lack of blinding may have reduced motivation in the control communities. However, the number of households lost during follow-up and the number of days under observation were almost identical in both arms. Additionally, the control communities knew that they would receive the intervention after study end. Finally, a reduction of diarrhoea frequency of 20% might be insufficient to be well perceived, i.e. have a noticeable impact in a population with a high burden of child diarrhoea and will, thus, not result in a sustainable behavioural change. Faecal contamination in about 60% of the yards indicates a highly contaminated environment with presumably a large potential for transmission pathways other than consuming contaminated water. This simultaneous exposure to a multiplicity of transmission pathways may explain why we found no significant diarrhoea reduction due to SODIS.

On the other hand, our result of a 19% reduction in diarrhoeal episodes appears to be roughly consistent with results of the two other SODIS trials both from Maasai cultural settings conducted by Conroy and colleagues among children under 6 and 5-

16 years of age. They report a 16% reduction (in <6 years olds, two-weeks prevalence of 48.8% in intervention and 58.1% in control group) [8] and a 10.3% reduction in the two-weeks diarrhoea prevalence (in 5-16 year olds) [7]. However, these randomized controlled trials were undertaken in a Maasai socio-cultural setting assuring a 100% compliance (as stated by the authors) in water treatment behaviour through social control by Maasai elder who promoted the method [7,8]. In the results presented in these studies adjusted models with post-hoc selected covariates were presented (i.e. no unadjusted models were provided). These trials were carried out in conditions of heavily contaminated drinking water and very high diarrhoea rates,- important considerations when attempting to generalize these results. The only other – quasi randomized – trial to estimate the effect of solar water disinfection was carried out in the urban slum in Vellore and resulted in a remarkable reduction of diarrhoea among children <5 (incidence rate ratio: 0.64, 95% CI 0.48-0.86) despite 86% of SODIS-users drinking also untreated water [9].

To our knowledge this is the first community-randomized trial and the largest study so far to assess the effectiveness of the SODIS-method under typical social and environmental conditions in a general rural population setting where children drink untreated water.

Our study was sufficiently powered to detect a 33% reduction in the effectiveness of the SODIS-intervention and we accounted for clustered design in our analysis. Based on a post-hoc sample size calculations using the model-based estimate for the between-cluster variability (CV_c=0.27) we would have needed a study 2.5 times larger for a 20% difference to be significant.

The implementing NGO with a worldwide experience to disseminate SODIS adapted a campaign to local and cultural needs and also involved the public health and educational system in the roll-out. This comprehensive SODIS-campaign resulted in a mean SODIS usage of 32% on any given study day. In using the SODIS-use indicator based on the personal judgement of community-based staff we intended to measure actual use in combining objective, visible signs of use (e.g. bottles exposed to sunlight)

with proxies more responsive to actual treatment behaviour (e.g. SODIS-water can be offered to drink upon request). We consider this a restrictive, more conservative definition of SODIS-use compared to that in other studies which recorded reported use [9] or the number of bottles exposed to sunlight [36]. Both are indicators that can easily and reliably be measured but which are prone to over-reporting due to low specificity for actual use. Further studies will need to validate different compliance indicators and formally assess the dimension of reporting bias.

It is possible that respondents would like to please field staff and over-report use out of courtesy. Also, observing exposed bottles on the roof may overestimate use (Figure II.3) as some households anecdotally were noted to place bottles on the roof to avoid discussions with the SODIS-implementing NGO-staff. Figure II.3 is indicative of this phenomenon, as reported use at the beginning and reported use and satisfaction with the method at end of study reached the 80% mark – a usage figure consistent with other studies relying on reported compliance [9] and evaluation reports from grey literature. We conclude that self-reported SODIS-use may overestimate compliance and a combination of reported and objectively measurable indicators provides more accurate SODIS-compliance data.

There are limitations to our study. As in other studies [24,37], we observed a decline in the reporting of child diarrhoea during the observational period in both arms (Figure II.2). If true, seasonal variation of diarrhoea could be one possible cause, increased awareness leading to more attention to basic hygiene and hence to illness reduction may be another reason. Alternatively, the pattern could be due to survey fatigue.

Despite a comprehensive and intensive intervention promotion campaign, we detected no strong evidence for a significant reduction in the incidence rate of diarrhoea in children <5 years in families using SODIS in our trial in a typical setting in rural Bolivia. We believe that a clearer understanding of the discrepancy between laboratory and field results (obtained under typical environmental and cultural conditions), the role of compliance in effectiveness, and a direct comparison of

SODIS to alternate drinking water treatment methods is needed before further global promotion of SODIS.

Acknowledgments:

The authors greatly acknowledge the families who participated in the study, the dedication of our study communities and the support of the authorities of Cochabamba and Totora District that made this project possible. Project Concern International (PCI) allowed embedding this evaluation trial in their local community development plans, - we specifically thank the PCI implementation team around Carlos Morante and Luciano Cespedes and their field staff for their relentless efforts to intertwine research and development approaches. We are grateful to the study team: Roy Cordova (administrator Bolivia site), Freddy Arauco (data entry, supervisor), Abrahan Cuevas, Fernando Salvatierra, David Villaroel, Dora Claros, Elmer Garvizu, Alfonso Claure (field supervisors and sample collectors), Claudia Lazarte (medical practitioner) and the entire MMS team (Morbidity Monitoring Staff). We thank Sonia Peredes, Gabriela Almanza and Gonzalo Fillips (laboratory staff), Jenny Rochas, Edgar Sejas, Ana Maria Romero and Mirjam Mäusezahl for their personal and institutional support. Profs Lee Riley (UCB), Alan Hubbard (UCB), and Joseph Eisenberg, University of Michigan and formerly at UCB thankfully contributed to develop the research plan for this trial. We much appreciate the manifold administrative support of Catherine Wright (UCB) and Ulrich Wasser (STI, Basel). Marcel Tanner (STI) kindly reviewed the manuscript. Markus Niggli, Tim Haley, Michael Hobbins, Stephan Indergand contributed to the study implementation and initial analyses. The main financial support was from the National Institutes of Health (NIH), Award Number R01AI50087-01. In kind and financial support were also provided by Project Concern International, University of San Simon, SODIS Foundation, Cochabamba and the Swiss Tropical Institute, Basel. The contents of the paper are solely the responsibility of the authors and do not necessarily represent the official view of NIH. Portions of this manuscript were presented at the 12th International Congress on Infectious Diseases in Lisbon, Portugal, on June 16, 2006.

Author Contributions:

The principal investigators Drs Mäusezahl and Colford had full access to the data and take responsibility for the integrity of the data and accuracy of the data analysis.

Study concept and design: Mäusezahl, Colford

Acquisition of data: Christen, Cevallos, Zapata, Iriarte

Implementation of public health interventions: Alvarez Tellez, Christen, Cevallos, Daigl Cattaneo

Analysis and interpretation of data: Mäusezahl, Christen, Duran Pacheco, Hattendorf, Smith, Arnold, Colford

Drafting of the manuscript: Mäusezahl, Christen, Duran Pacheco, Hattendorf, Colford

Critical revision of the manuscript for important intellectual content: Mäusezahl, Christen, Duran Pacheco, Hattendorf, Colford, Arnold

Statistical analysis: Duran Pacheco, Hattendorf, Daigl Cattaneo, Arnold

Obtained funding: Mäusezahl, Colford

Administrative, technical, or material support: Mäusezahl, Christen, Cevallos, Daigl Cattaneo, Hattendorf

Study coordination and supervision: Christen, Cevallos, Mäusezahl

Potential conflicts of interest of the authors: none declared

Role of the Sponsor:

The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

References:

1. WHO (2005) The World Health Report 2005 - make every mother and child count.

2. Pruss A, Kay D, Fewtrell L, Bartram J (2002) Estimating the burden of disease from water, sanitation, and hygiene at a global level. Environ Health Perspect 110: 537-542.

3. Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, Colford JM, Jr. (2005) Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. Lancet Infect Dis 5: 42-52.

4. Zwane AP, Kremer M (2007) What Works in Fighting Diarrheal Diseases in Developing Countries? A Critical Review. National Bureau of Economic Research Working Paper Series No. 12987

5. Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S (2007) Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. BMJ 334: 782.

6. Sobsey M (2002) Managing Water in the home: Accelerated Health Gains from Improved water. WHO/SDE/WSH/02.07; The World Health Organization: Geneva.

7. Conroy RM, Elmore-Meegan M, Joyce T, McGuigan KG, Barnes J (1996) Solar disinfection of drinking water and diarrhoea in Maasai children: a controlled field trial. Lancet 348: 1695-1697.

8. Conroy RM, Meegan ME, Joyce T, McGuigan K, Barnes J (1999) Solar disinfection of water reduces diarrhoeal disease: an update. Arch Dis Child 81: 337-338.

9. Rose A, Roy S, Abraham V, Holmgren G, George K, Balraj V, Abraham S, Muliyil J, Joseph A, Kang G (2006) Solar disinfection of water for diarrhoeal prevention in southern India. Arch Dis Child 91: 139-141.

10. Eisenberg JN, Scott JC, Porco T (2007) Integrating disease control strategies: balancing water sanitation and hygiene interventions to reduce diarrheal disease burden. Am J Public Health 97: 846-852.

11. McCuin RM, Bukhari Z, Sobrinho J, Clancy JL (2001) Recovery of Cryptosporidium oocysts and Giardia cysts from source water concentrates using immunomagnetic separation. J Microbiol Methods 45: 69-76.

12. Murray DM (1998) Design and analysis of group-randomized trials. New York, Oxford: Oxford University Press.

13. Edwards SJ, Braunholtz DA, Lilford RJ, Stevens AJ (1999) Ethical issues in the design and conduct of cluster randomised controlled trials. BMJ 318: 1407-1409.

14. Chingono A, Lane T, Chitumba A, Kulich M, Morin S (2008) Balancing science and community concerns in resource-limited settings: Project Accept in rural Zimbabwe. Clin Trials 5: 273-276.

15. Ranson MK, Sinha T, Morris SS, Mills AJ (2006) CRTs--cluster randomized trials or "courting real troubles": challenges of running a CRT in rural Gujarat, India. Can J Public Health 97: 72-75.

16. Hayes RJ, Bennett S (1999) Simple sample size calculation for cluster-randomized trials. Int J Epidemiol 28: 319-326.

17. Ministry of Health, Bolivia. Situación de Salud Bolivia 2004,. La Paz/Bolivia. Available: http://www.sns.gov.bo/snis/.

18. Clasen T, Roberts I, Rabie T, Schmidt W, Cairncross S (2006) Intervention to improve water quality for preventing diarrhoea. Cochrane Database of Systematic Reviews, 3(3):CD004794.

19. WHO (1988) Persistent diarrhoea in children in developing countries: memorandum from a WHO meeting. Bull World Health Organ 66: 709-717.

20. Baqui AH, Black RE, Yunus M, Hoque AR, Chowdhury HR, Sack RB (1991) Methodological issues in diarrhoeal diseases epidemiology: definition of diarrhoeal episodes. Int J Epidemiol 20: 1057-1063.

21. Morris SS, Cousens SN, Lanata CF, Kirkwood BR (1994) Diarrhoea--defining the episode. Int J Epidemiol 23: 617-623.

22. Wright JA, Gundry SW, Conroy R, Wood D, Du PM, Ferro-Luzzi A, Genthe B, Kirimi M, Moyo S, Mutisi C, Ndamba J, Potgieter N (2006) Defining episodes of diarrhoea: results from a three-country study in Sub-Saharan Africa. J Health Popul Nutr 24: 8-16.

23. Morris SS, Cousens SN, Kirkwood BR, Arthur P, Ross DA (1996) Is prevalence of diarrhea a better predictor of subsequent mortality and weight gain than diarrhea incidence? Am J Epidemiol 144: 582-588.

24. Luby SP, Agboatwalla M, Painter J, Altaf A, Billhimer W, Keswick B, Hoekstra RM (2006) Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. Trop Med Int Health 11: 479-489.

25. Hobbins MA (2004) Home-based drinking water purification through sunlight: from promotion to health effectiveness [dissertation]. Basel, Switzerland: Swiss Tropical Institute, University Basel.

26. Young ML, Preisser JS, Qaqish BF, Wolfson M (2007) Comparison of subject-specific and population averaged models for count data from cluster-unit intervention trials. Stat Methods Med Res 16: 167-184.

27. Twisk JWR (2006) Applied Multilevel Analysis: A Practical Guide for Medical Researchers. UK: Cambridge University Press.

28. Turner RM, Omar RZ, Thompson SG (2006) Constructing intervals for the intracluster correlation coefficient using Bayesian modelling, and application in cluster randomized trials. Stat Med 25: 1443-1456.

29. Sommer B, Mariño A, Solarte Y, Salas ML, Dierolf C, Valiente C, Mora D, Rechsteiner R, Setter P, Wirojanagud W, Ajarmeh H, Al-Hassan A,

Wegelin M (1997) SODIS - an emerging water treatment process. J Water SRT 46: 127-137.

30. Campbell MK, Elbourne DR, Altman DG (2004) CONSORT statement: extension to cluster randomised trials. BMJ 328: 702-708.

31. Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S (2007) Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. BMJ 334: 782.

32. Kumar V, Mohanty S, Kumar A, Misra RP, Santosham M, Awasthi S, Baqui AH, Singh P, Singh V, Ahuja RC, Singh JV, Malik GK, Ahmed S, Black RE, Bhandari M, Darmstadt GL (2008) Effect of community-based behaviour change management on neonatal mortality in Shivgarh, Uttar Pradesh, India: a cluster-randomised controlled trial. Lancet 372: 1151-1162.

33. Campbell R, Starkey F, Holliday J, Audrey S, Bloor M, Parry-Langdon N, Hughes R, Moore L (2008) An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial. Lancet 371: 1595-1602.

34. Schmidt WP, Cairncross S (2009) Household Water Treatment in Poor Populations: Is There Enough Evidence for Scaling up Now? Environ Sci Technol. 2009 Feb 15;43(4):986-92..

35. Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S (2007) Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. BMJ 334: 782.

36. SODIS homepage. http://www.sodis.ch.

37. Colford JM, Jr., Wade TJ, Sandhu SK, Wright CC, Lee S, Shaw S, Fox K, Burns S, Benker A, Brookhart MA, van der LM, Levy DA (2005) A randomized, controlled trial of in-home drinking water intervention to reduce gastrointestinal illness. Am J Epidemiol 161: 472-482.

Chapter III:

Performance of analytical methods for overdispersed counts in cluster randomized trials: sample size, degree of clustering and imbalance

Gonzalo Durán Pacheco[1], Jan Hattendorf[1], John M. Colford, Jr.[2], Daniel Mäusezahl[1], Thomas Smith[3]

[1] *Department of Public Health and Epidemiology, Interventions and Health Systems Unit, Swiss Tropical Institute*
[2] *Division of Epidemiology, School of Public Health, University of California, Berkeley (UCB), California*
[3] *Department of Public Health and Epidemiology, Biostatistics and Epidemiology Unit, Swiss Tropical Institute, P.O. box 4002 Basel, Switzerland.*

This paper has been published in: *Statistics in Medicine* (2009); 28: 2989–3011

SUMMARY

Many different methods have been proposed for the analysis of cluster randomized trials (CRTs) over the last 30 years. However, the evaluation of methods on overdispersed count data has been based mostly on the comparison of results using empiric data; i.e. when the true model parameters are not known. In this study, we assess via simulation the performance of five methods for the analysis of counts in situations similar to real community-intervention trials. We used the Negative Binomial distribution to simulate overdispersed counts of CRTs with two study arms, allowing the period of time under observation to vary among individuals. We assessed different sample sizes, degrees of clustering and degrees of cluster-size imbalance. The compared methods are: (i) The two-sample *t* test of cluster-level rates, (ii) Generalized estimating equations (GEE) with empirical covariance estimators (iii) GEE with model-based covariance estimators, (iv) Generalized Linear Mixed Models (GLMM) and (v) Bayesian Hierarchical Models (Bayes-HM). Variation in sample size and clustering led to differences between the methods in terms of coverage, significance, power and random-effects estimation. GLMM and Bayes-HM performed better in general with Bayes-HM producing less dispersed results for random-effects estimates although upward biased when clustering was low. GEE showed higher power but anticonservative coverage and elevated type I error rates. Imbalance affected the overall performance of the cluster-level t-test and the GEE's coverage in small samples. Important effects arising from accounting for overdispersion are illustrated through the analysis of a community-intervention trial on Solar Water Disinfection in rural Bolivia.

Keywords: Negative Binomial count data, community-cluster randomized trials, GLMM, GEE, Bayesian hierarchical models, t-test.

1. INTRODUCTION

Cluster Randomized Trials (CRTs) are studies for which the unit of random allocation is a group of individuals rather than an individual. The cluster units might be well defined geographical areas, communities, schools, hospitals, worksites, etc., and the reasons for assigning entire groups to the intervention range from logistical convenience to the impossibility of operating/delivering the intervention at individual level [1].

As health outcomes are measured at individual level, and individuals are likely to be correlated within a cluster, statistical analysis of such trials without allowance for clustering might produce inflated type I error rates in statistical testing and falsely narrow confidence intervals. For these reasons attention has been given to the development and study of statistical methods that address within-cluster dependence over the past 30 years [2, 3].

A number of methods have been proposed for the analysis of different outcomes. The analysis of cluster-level summary statistics (rates, odds or means) by the basic t-test, Wilcoxon's U-test, Chi-square tests, etc., is well described [4-6]. In addition, a more extensive class of statistical models including the multilevel, hierarchical or random effect regression models, more broadly typified as Generalized Linear Mixed Models (GLMM), has been developed in parallel to the Generalized Estimating Equation (GEE) methods, to estimate the effect of covariates while allowing for intracluster correlation [7-10]. These methods can be divided into two main classes: the conditional or Cluster Specific (CS) and the marginal or Population Averaged (PA) models with GLMM and GEE respectively as the prominent representatives. The main distinction between CS and PA models is whether the regression coefficients describe a cluster level or the average population response to the covariates' changes. A secondary distinction is in the nature of the assumed within-cluster dependence. CS models condition the model on random effects which reflect the correlation among observations of the same cluster while GEE account for correlation by incorporating predefined correlation structures to describe the nature of within-clusters dependencies [11]. Alternatively, the Bayesian paradigm proposes highly flexible

methods to analyse random-effects models overcoming the computational problems of GLMM and providing a full distributional answer to the estimate values of the parameters [12-14].

Previous research has concentrated on the performance of such methods in the context of CRTs for continuous and binary data. Theoretical equivalences and a comprehensive assessment through simulation are available for these outcomes [15-23]. Some attention has been also focused on the analysis of counts and incidence rates, although a thorough evaluation particularly in the context of clustered count data under overdispersion, have been generally done by means of illustrations in the form of analysis of real datasets where the true model parameters were unknown [6, 24-28]. The number of clusters and the degree of clustering appear among the factors that greatly affect the performance of the methods, and are considered in the planning of new CRT. Although equal cluster sizes may be assumed in the design, balance is rarely found after data collection, and such imbalance is known to affect the analysis of binary data[29-32].

In the present study we assess the statistical performance of 5 methods for analysing CRTs by simulating situations close to real community-randomized trials, when a count outcome, observed in individuals with different follow-up periods, is overdispersed. The number of clusters and cluster size imbalance are assessed across a gradient of intercluster variability. The methods compared are: (i) the two-sample *t* test, (ii) GEE with empirical covariance estimator, (iii) GEE with model-based covariance estimator, (iv) GLMM and (v) the Bayesian Hierarchical Models. We illustrate the results with the motivating example of a CRT of solar water disinfection in rural Bolivia.

2. A MOTIVATING EXAMPLE

Solar drinking water disinfection (SODIS) is a low-cost, point-of-use water purification method that uses solar energy to inactivate waterborne pathogens. The

combined effect of UV-A radiation and the increase of water temperature has been shown to be efficacious in inactivating microbiological pathogens, when water is exposed to sunlight in plastic bottles [33, 34]. However, there has been limited evidence of its effectiveness at reducing the burden of waterborne diseases in populations consuming contaminated water.

A community randomized trial (BoliviaWET) was conducted to evaluate the effect of SODIS promotion in reducing diarrhoea among children under 5 years of age [35]. The study took place in 22 rural communities of the Cochabamba department in Bolivia. The communities were pair-matched by community diarrhoeal incidence at baseline, and the SODIS intervention was randomly assigned to one community within each pair. The intervention was implemented through 15 months of intensive promotion of the SODIS-method along with personal and home-hygiene educational training in the intervention communities.

Diarrhoea, was monitored by a surveillance monitoring system for one year, and individual diarrhoea occurrence was assessed daily. In this paper we analyze the effects of the intervention on the primary outcome, i.e. the incidence rate expressed as the number of episodes per child (Y) per time at risk (t) without considering potential confounders (child age, sex, hand washing habits) and ignoring pair-matching. We henceforth use the data of the trial for illustration purposes only.

3. ANALYTICAL METHODS

3.1. The t-test

Consider a two-arm CRT with a count outcome Y_{ijl} (values = 0,1,2,…) observed in a time period t_{ijl}, on the individual l (l =1,…n_{ij}), from cluster j (j=1,…,k_i), receiving the intervention i (i=1,2). The analysis considers the cluster-level rates r_{ij} of the counts Y_{ijl} per observed time t_{ijl} as the units for the analysis.

$$r_{ij} = \sum_{l=1}^{n_{ij}} y_{ijl} \Big/ \sum_{l=1}^{n_{ij}} t_{ijl} \ . \tag{1}$$

Defining the mean rates in the i^{th} arm by $\bar{r}_{i.} = \frac{1}{k_i}\sum_{j=1}^{k_i} r_{ij}$, the effect of the intervention can be estimated by the ratio of the group mean rates:

$$RR = \frac{\bar{r}_{1.}}{\bar{r}_{2.}}, \tag{2}$$

known as the Rate Ratio (RR). By a Taylor series approximation, the 95% confidence intervals (CI) are calculated as exp[log RR $\pm$ $t_{k1 + k2 - 2,\ 0.025}\sqrt{V}$] [6], with

$$V = V(\log RR) \approx \frac{s_1^2}{k_1 \bar{r}_{1.}^2} + \frac{s_2^2}{k_2 \bar{r}_{2.}^2} \tag{3}$$

When (2) is used as a point estimate, hypothesis testing can be performed through an unpaired t-test on the cluster rates [6] as follows:

$$T = \frac{\bar{r}_{1.} - \bar{r}_{2.}}{S\sqrt{(1/k_1 + 1/k_2)}} \sim t_{K-2;\alpha}\,, \tag{4}$$

$$S = \sqrt{\frac{(k_1 - 1)s_1^2 + (k_2 - 1)s_2^2}{K - 2}} \quad \text{and} \quad s_i^2 = \frac{1}{k_i - 1}\sum_{l=1}^{k_i}(r_{ij} - \bar{r}_{i.})^2 \,,$$

where $K = k_1 + k_2$ is the total number of clusters. T follows a *Student* distribution with $K - 2$ degrees of freedom when the r_{ij} are normally distributed, but this normality assumption is not usually met in CRTs. Nevertheless, since simulations have shown that the t-test is robust to the violation of the underlying assumptions [36] this may be a reasonable analytical approach. A test on the rate ratio (2) using the Taylor's series approximation in (3) could be also performed, however (4) is much easier to implement and produces similar results.

3.2. Random-effect models.

A more complete representation of the structure of the data is given by specifying a Generalized Linear Mixed Model (GLMM). GLMM represents an extension of generalized linear models (GLM) specified by a linear predictor, link function, variance function and outcome distribution at the cluster and individual levels. The linear predictor can be specified as follows:

$$\eta_{jl} = \mathbf{x}'_{jl}\boldsymbol{\beta} + \mathbf{z}'_{jl}\boldsymbol{\nu}_j, \qquad (5)$$

where $\mathbf{x}_{jl}$ is the vector of covariates observed on individual l nested within the cluster j, $\boldsymbol{\beta}$ is the vector of fixed-effects regression parameters, $\mathbf{z}_{jl}$ the vector of variables having random effects, and $\boldsymbol{\nu}_j$ the vector of random effects which are usually assumed to follow a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$.

The link function $g(.)$, relates the expected value or mean μ_{jl} of the outcome variable Y_{jl} (i.e. $E[Y_{jl}] = \mu_{jl}$) to the linear predictor η_{jl}, i.e.:

$$g(\mu_{jl}) = \eta_{jl}.$$

The variance can be specified in terms of the mean μ_{jl}, as $V(Y_{jl}) = \phi v(\mu_{jl})$, where ϕ is called the overdispersion parameter. The later two specifications depend on the distribution of the outcome Y_{jl} which is assumed to fall within the exponential family of distributions [37].

The expected value of the outcome variable in terms of the linear predictor (via the link function) is then:

$$\mu_{jl} = E\left[Y_{jl} \mid \mathbf{x}_{jl}, \boldsymbol{\nu}_j\right], \qquad (6)$$

and represents the expectation of the conditional distribution of the outcome given the random effects. As a consequence GLMM are referred to as conditional models in contrast to GEE which are considered methods to estimate marginal effects.

Considering the random-intercepts model with the count outcome Y_{jl} (i.e. 0,1,2,…) of a two-arm CRT, the linear predictor of the expected number of counts μ_{jl} has the following form:

$$\log(\mu_{jl}) = \eta_{jl} = \beta_0 + \beta x_j + \nu_j , \quad (7)$$

where the link $g(.)$ is the log function that transforms the scale of the counts (permitting only positive values) to the scale of the linear predictor η_{jl} which can take any value in the real line; β_0 is the intercept, β the log of the RR of the intervention x_j (0 = control, 1 = intervention) implemented in cluster j and ν_j is the random effect of the j^{th} cluster ~ $N(0,\sigma^2_c)$. If the time over which the counts were observed differs among individuals, being t_{jl} the time of observation of individual l in cluster j, the linear predictor is augmented as

$$\log(\mu_{jl}) = \log(t_{jl}) + \beta_0 + \beta x_j + \nu_j , \quad (8)$$

also expressed as $\mu_{jl} / t_{jl} = \exp(\beta_0 + \beta x_j + \nu_j)$ to reflect that it is the number of counts per follow-up period that is modelled. The term $\log(t_{jl})$ is often called the offset.

We consider two distributional assumptions for count data:

1) Poisson distributed counts, Y_{jl} ~ Poi(μ_{jl}), with variance function $V(Y_{jl}) = \phi v(\mu_{jl}) = \mu_{jl}$ where ϕ is assumed to be 1; i.e. the mean equals the variance or equidispersion, property that is rarely found in real practice.

2) Negative Binomial (NB) distributed counts, Y_{jl} ~ NB(s, μ_{jl}) with a variance function $V(Y_{jl}) = \phi v(\mu_{jl}) = \phi(\mu_{jl} + s\mu_{jl}^2)$, where ϕ is assumed to be 1 and s is the NB

overdispersion parameter, indicating that the NB distribution models overdispersion implicitly by its parameter *s* [38].

We consider two alternative approaches for parameter estimation of random-effect models:

(i) *Maximum-likelihood based methods*. To estimate the model parameters, the solution of integrals of the likelihood function over the random-effects is needed but can be numerically intensive particularly for discrete data where solutions may not have a closed form. Taylor's series (linearizations) approximations [39] as well as numerical integration [40] for evaluating such integrals have been proposed. We apply the first class, specifically Restricted Pseudo Likelihood estimation as implemented in the GLIMMIX procedure in SAS v9 [41] and denote it henceforth as GLMM.

(ii) *Bayesian estimation via a Markov chain Monte Carlo algorithm*. In the Bayesian framework, the computation of posterior probabilities P(θ | *data*) is the focal concern. For a CRT with count outcome Y_{jl} ~ Poi(μ_{jl}) or Y_{jl} ~ NB(s, μ_{jl}) and a model log(μ_{jl}) = log(t_{jl}) + β_0 + βx_j + ν_j; ν_j ~ N(0, σ_c^2) the posterior probabilities P(β_0, β, σ_c^2,s | Y) are calculated by updating the likelihood $f(Y; \beta_0, \beta, \sigma_c^2, s)$ with the prior P(β_0, β, σ_c^2,s) as established by the Bayes' principle, using Markov chain Monte Carlo simulation (MCMC), Gibbs sampling specifically as defined in the WinBugs Software v1.4 [13, 42].

3.3. Generalized Estimating Equations (GEE).

GEE are useful to estimate marginal or PA effects in the context of correlated data. As an extension to GLM, GEE is applicable to different types of outcomes by defining a link function $g(.)$, a linear predictor ($\eta_{jl} = \mathbf{x}'_{jl}\boldsymbol{\beta}$), a variance function $v(\mu_{jl})$, and a working correlation matrix that is typically assumed to be the same across all clusters [43]. Unlike GLMM, in GEE no distributional assumptions are made on Y_{jl} and inferences are asymptotically unbiased and efficient as long as the mean and variance

functions are correctly characterized [44]. As the linear predictor η_{jl} does not depend on any random effect, the mean response reflects the average effect of the population. On the other hand, the variance of Y_{jl} depends on $v(\mu_{jl})$ and $\mathbf{R}(\alpha)$, the working correlation matrix. For more details and contrasts with GLMM we refer the reader to Zeger et al, 1988, Young, 2007 and Fitzmaurice 2004 [11, 27, 45].

Different types of correlation structures have been proposed for $\mathbf{R}(\alpha)$: *Independence* where $\mathbf{R}(\alpha)$ is an identity matrix, i.e. individuals are all independent. *Exchangeable* where $\mathbf{R}(\alpha)$ is a matrix with 1s in the diagonal and α elsewhere. Note that α, the correlation of individuals within the same cluster, is assumed to be constant across clusters. *Unstructured* where $\mathbf{R}(\alpha)$ is a symmetric matrix with 1s in the diagonal and $\alpha_{ll'}$ elsewhere. Other structures are also proposed [7, 43, 46].

Assuming a two-arm CRT with a count outcome Y_{jl} per follow-up time t_{jl} , the marginal model is

$$\log\left(\mu_{jl}\right) = \log\left(t_{jl}\right) + \beta_0^{PA} + \beta^{PA} x_j \qquad (9)$$

with the same characterizations of models (7) and (8). The intervention effect is labelled differently to make clear the PA interpretation of the marginal model (9) in contrast to the CS interpretation in models (7) and (8). The expectance $\mu_{jl} = \mathrm{E}[Y_{jl} \mid x_{jl}]$ contrasts to that of (6). The log link is complemented with the variance functions $\phi v(\mu_{jl}) = \mu_{jl}$ or $\phi v(\mu_{jl}) = \phi\left(\mu_{jl} + s\mu_{jl}^2\right)$, similar to the Poisson or NB GLMMs respectively. Note however that the β coefficient of GLMM has both CS and PA interpretations when the log link is used [27].

All the parameters are estimated by solving the estimating equations:

$$\mathbf{U} = \mathbf{D}'\mathbf{V}^{-1}\left(\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\beta})\right)$$

where **D** contains the partial derivatives $\delta\mu/\delta\beta$, **V** contains $\phi v(\mu_{jl})$ and **R**(α), and finally **Y** and **μ**(β) are the vectors of observations and mean functions respectively. We consider two alternative standard errors estimating methods:

(i) *Empirical covariance estimates*. If **R**(α) is incorrectly specified the variance of the outcome is inefficient providing inaccurate standard errors for the **β** estimators. This problem can be overcome by using the "sandwich" or "robust" variance estimator (empirical estimator), popularized by Liang & Zeger [44] which is consistent for large sample sizes even when **R**(α) is incorrectly specified under the assumption of missing at random. However, it was shown to perform poorly for small sample sizes [2].

(ii) *Model-Based covariance estimates.* If **R**(α) is correctly specified the inverse of the Fisher information matrix also known as the model-based estimator, can be used as an estimator of the covariance of **β,** producing consistent standard errors even in scenarios with small number of clusters [27, 43].

3.4. Simulations

Datasets were generated for different number of clusters (K = 10, 20, 40), degrees of imbalance (balanced, slightly and highly imbalance designs) and degree of clustering (σ_c = 0.05, 0.15 and 0.40 as the between-cluster standard deviation on the log risk scale). The number of individuals per cluster was set at 30 for balanced designs, while for slightly and highly imbalanced designs the cluster size was generated from normal distributions with mean 30 and s.d.=6 and s.d.=18 individuals per cluster (c.v.=20% and 60%) respectively. The fractional cluster sizes were rounded up to the closest integer and the number of individuals per cluster was truncated to a minimum of 8, assuming 8 to be too small for community recruitment in large field trials (e.g. min of the motivating example was 24). A different exposure time t_{jl} per individual was assumed, with t_{jl} being sampled from a negative skewed distribution similar to the one observed in the motivating example: skewness -1.4, mean 290 and s.d. 100, through a power transformation: $t_{jl} = 80(x_{jl}^{1/4})$ where $x \sim N(200,100)$. The control-group event rate θ was set at 5/365 (events per days at risk), and a protective efficacy of 30% was assumed implying a RR of $\exp(\beta) = 0.70$. A null effect was also simulated in order to

assess the significance level. A cluster effect δ_j was set to act multiplicatively on the mean and whose logarithm was normally distributed with mean 0 and s.d. = σ_c. Note that σ_c under the log link and by a Taylor series expansion is approximately equal to the between-cluster coefficient of variation (CV_c) [6, 47]. To simulate the within cluster variation and overdispersion specifically, the number of events Y_{jl} were produced from a NB distribution $Y_{jl} \sim NB(\mu_{jl}, s)$, with mean $\mu_{jl} = \theta t_{jl} \delta_j$ and $\mu_{jl} = \theta t_{jl} \delta_j \exp(\beta)$ for control and intervention clusters respectively, variance $v(\mu) = \mu + s\mu^2$ and a fixed overdispersion of $s = 0.5$.

One thousand datasets were produced using different seeds for each of the $3 \times 3 \times 3$ possible arrangements. Each dataset was subsequently analysed by: i) The t-test of cluster-level rates as defined in (1) – (4), ii) GEE with empirical covariance estimators (GEE-Emp), iii) GEE with model-based covariance estimators (GEE-MB), both implemented in SAS v9 by the GENMOD procedure [48] specified according to model (9), with a log link, a NB variance function and an exchangeable correlation; iv) GLMM as implemented in the GLIMMIX procedure of SAS v9 [41] based on model (8) assuming a NB distribution.

A random subset of 300 datasets were analysed using method v) a Bayesian hierarchical model (Bayes-HM) implemented in WinBugs v1.4., specified according to model (8) assuming NB distributed counts (the high computational demands precluded analysing all 1000 datasets by this method). For this analysis, uninformative priors were used : $\beta_0 \sim N(0, 10^6)$, $\beta \sim N(0, 10^6)$, $\sigma_c^2 \sim IG(0.001, 0.001)$, $s \sim \Gamma(0.001, 0.001)$ (IG=Inverse Gamma distribution). A SAS-WinBugs interface was written to analyse the replicate datasets per arrangement in SAS. The convergence was previously assessed in WinBugs by running two chains with dispersed initial values throughout the parameter space and comparing the between and within chain variation in sample datasets for each of the 27 situations. Convergence was achieved before 5,000 iterations, but 15,000, 10,000 and 7,000 iterations after 1,000 burn-in were implemented in the interface for $K = 10$, 20 and 40 respectively. The posterior 2.5% and 97.5% quantiles were reported as the intervals (CI for simplicity) and the median as the point estimate.

For each method, performance in point and interval estimation as well as hypothesis testing were assessed in terms of:

- Relative Bias as: |mean estimated RR – true RR|/true RR*100.
- The empirical standard errors (EmpSE), computed as the root square of the variance of the RR estimates across the simulated datasets.
- The width of the CI as the range between the upper and lower confidence limits.
- Coverage probability of the confidence interval (CI) expressed as the proportion of intervals that contained the true RR.
- Type I error rate, as the proportion of significant findings at 0.05 level when the true RR = 1.
- Statistical power as the proportion of significant results at 0.05 level when the true RR = 0.7.

Finally, estimation of the underlying between-cluster standard deviation σ_c was also assessed. For the t-test, the ANOVA variance component method was used

$$\hat{\sigma}^2_{clust} = \frac{MS_c - MS_e}{n_0} \qquad (10)$$

where MS_c is the intercluster mean squares, MS_e, the intracluster mean squares and n_0 a weighted mean cluster size (see Donner & Klar, 1994, Ukoumunne, 2002 for full details [4, 49]). Since σ_c is $\log(\mu_{jk})$ scaled and σ_{clust} is in the rate scale, the between-cluster coefficient of variation CV_c [47] was estimated by $\hat{\sigma}_{clust}/\bar{r}..$ allowing to compare the cluster variability of the ANOVA method with that of σ_c produced by GLMM and Bayes-HM. Indeed, by a Taylor's first order expansion of μ_{jk} around μ at the log link, σ_c is found to approximately equal CV_c. GEE correlation estimates were not considered.

4. RESULTS

Simulations

Bias and empirical standard error.

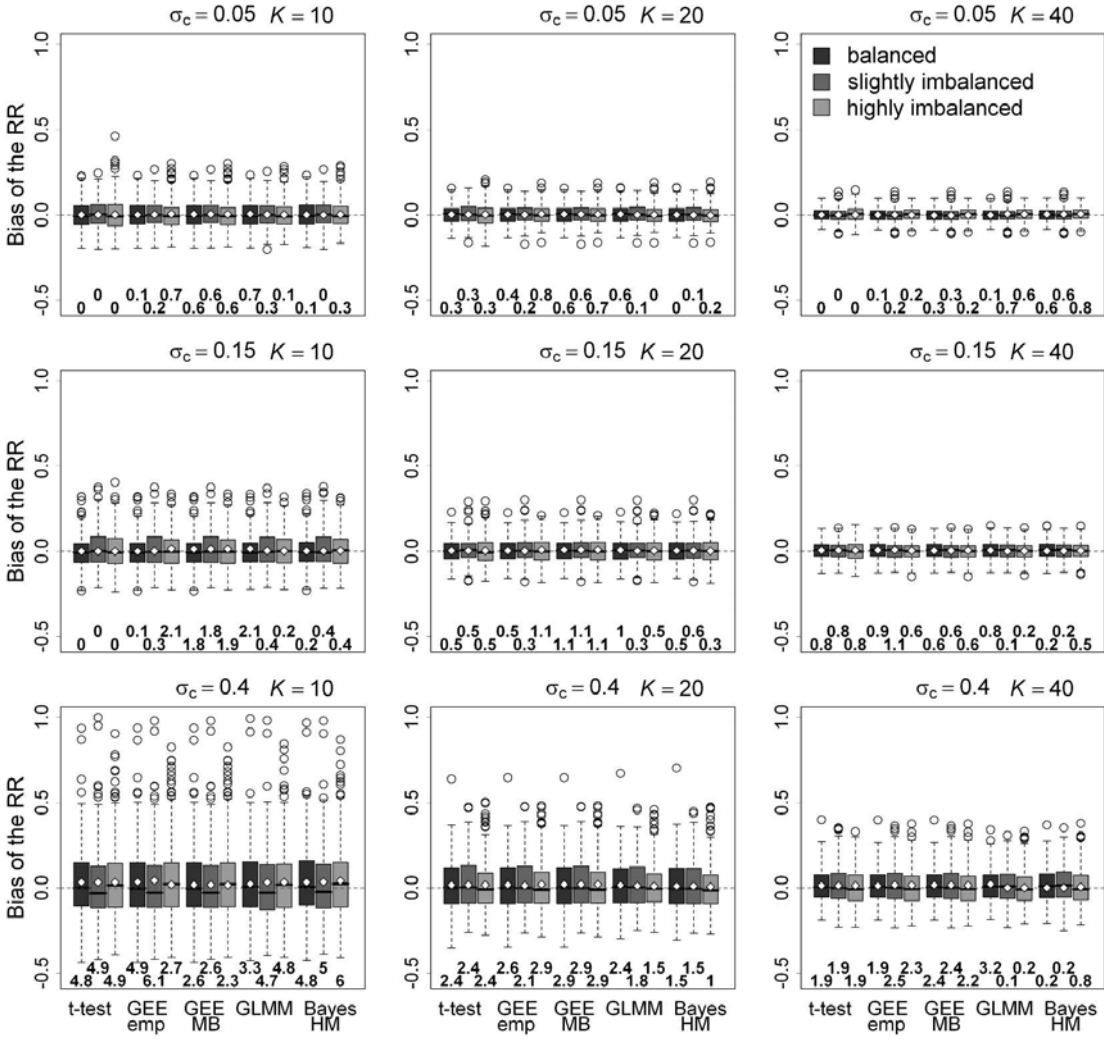
The distribution of the absolute bias of the RR estimates with respect to the true 0.70 value is depicted in Figure III.1. All the methods across the studied situations provided an average bias around 0, while the level of dispersion depended on K and σ_c. The analysis of the relative bias indicated that in 90% of the studied combinations, the bias was below 3%, with the highest values when σ_c = 0.40 and K = 10. The relative bias and EmpSE were more sensitive to K and σ_c than to the methods, although the advantage of large sample size on the relative bias was evident only when σ_c = 0.40. The EmpSE, tended to decline similarly in all methods with the increase of K, and increased with σ_c, No noteworthy differences in bias were found between: the t-test, GEE-Emp, GEE-MB, GLMM. Although Bayes-HM occasionally differed from the other methods, the differences were negligible compared to the ones due to K or σ_c.

Width of the confidence interval.

The average width of the CI and its coefficient of variation (c.v.) across replicate datasets are given in Table III.1. GEE-Emp and GEE-MB produced the narrowest intervals among the methods. Bayes-HM and GLMM yielded less variable interval widths across replicate datasets compared to the t-test, GEE-emp and GEE-MB. The degree of imbalance made no difference to the mean width of the CI except for the t-test which showed higher and more unstable widths under high imbalance. Imbalance affected however the stability of the CI widths of the other methods with more variable widths with higher imbalance. This effect was no longer evident when σ_c = 0.40, where high clustering appears to conceal the effect of high imbalance. As expected, the CIs were narrower with larger sample sizes, and wider with larger σ_c.

The between-datasets variability of the widths followed the same pattern; more stable widths were found with larger *K*, and larger c.v. were associated with larger σ_c.

Figure III.1: Distribution of the absolute bias of the relative rate (RR) of 5 statistical methods for overdispersed counts in cluster randomized trials at different: total number of clusters *K*, between-cluster variation σ_c and levels of cluster size imbalance.



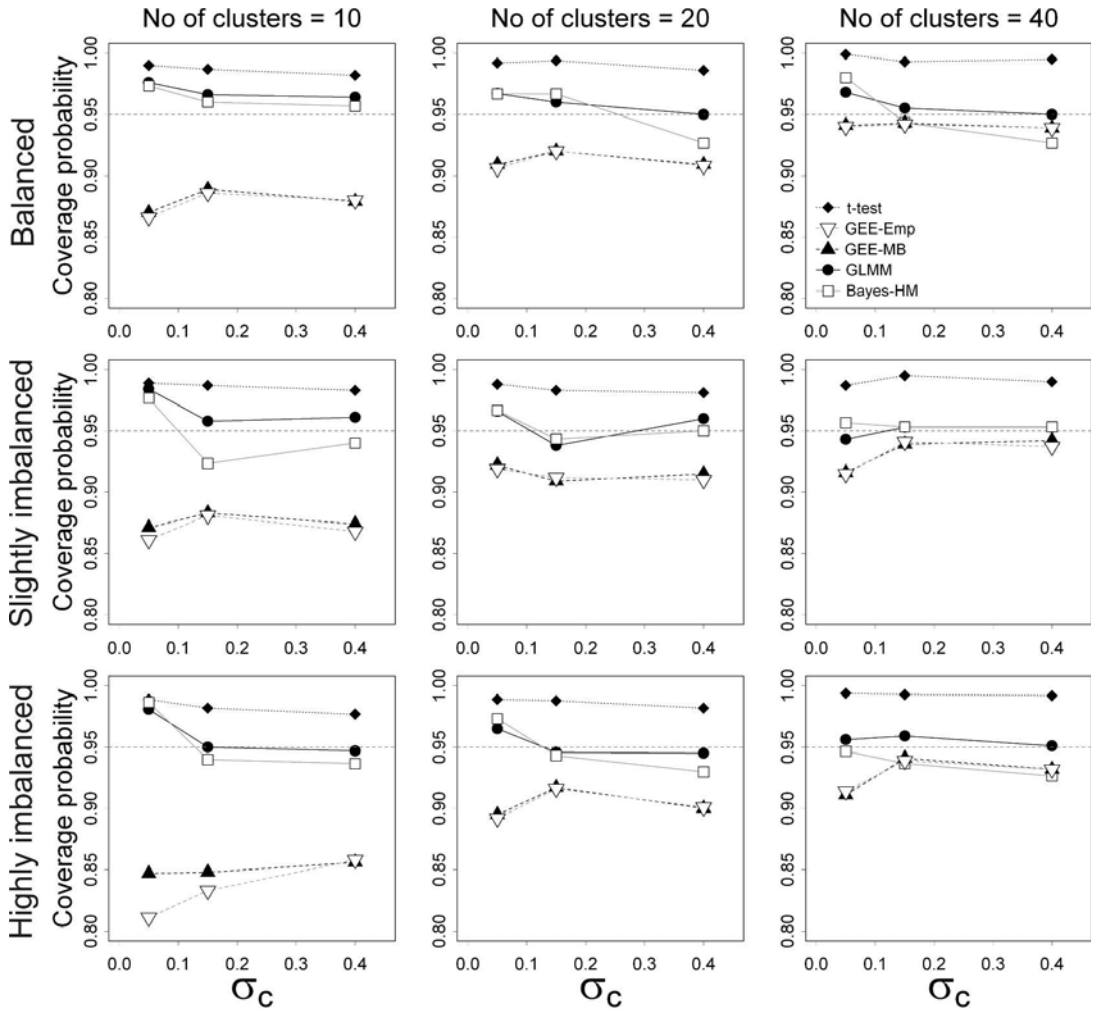
Footnote: Numbers at the bottom represent the relative bias (%).

Coverage Probability

Figure III.2 displays the coverage probabilities of the CIs for RRs obtained by the 5 methods at different *K*, σ_c and levels of imbalance. The CI coverage for GEE methods were in most cases lower than nominal, but approached the 95% reference when *K* = 40. The t-test, on the contrary, always provided higher than nominal coverage,

possibly because of the wide intervals resulting from the imprecision arising from the use of the cluster-level rates as analysis units. A cluster-level t-test weighted by cluster size would have been expected to produce more efficient results. GLMM and Bayes-HM produced coverage around the nominal in all the scenarios and somewhat higher than nominal when $\sigma_c = 0.05$. The differences between those two methods were due to the coverage proportions computed from different total number of replicates analysed (1000 vs 300). No difference was observed when the coverage proportions came form the same 300 replicates. Imbalance appeared to accentuate the unfavourable coverage of GEE methods specially when $K = 10$.

Figure III.2: Coverage provability of 5 analytical methods for overdispersed counts of clustered randomized trials at different: between-cluster variation σ_c, total number of clusters K and levels of cluster size imbalance.



Type I error rates and Power

The distinction between the GEE and the other methods is also evident in the type I error rates and the power (Table III.2). The risk of type I error was higher than nominal for both GEE-Emp and GEE-MB and in general high as compared to the t-test, GLMM and Bayes-HM. This risk nonetheless approached the nominal values when K increases and especially when $K = 40$ and $\sigma_c = 0.4$. GLMM and Bayes-HM had normally lower probabilities of detecting false significant results under all the studied conditions, except when $\sigma_c = 0.4$ where Bayes-HM yielded sometimes higher error rates than GLMM. These two methods produced conservative error rates when $\sigma_c = 0.05$ and $K \leq 20$. There was no marked difference by types of imbalance, nor was a clear relation with K or σ_c observed.

The analysis of statistical power refers to the power required to detect the simulated 30% protective reduction in the incidence rate. GEE-Emp and GEE-MB were generally more powerful than the other three methods, most clearly in the unfavourable situations: $K \leq 20$ and $\sigma_c = 0.4$. No clear differences in power were observed when comparing degrees of imbalance except for the t-test which showed a consistent decrease in power with higher imbalance. In addition to the effect of K, power was influenced by the degree of clustering, i.e. all the methods report rather high probabilities of detecting true significant effects when $\sigma_c = 0.05$ while regardless the sample size, the power of all methods falls below 80% when $\sigma_c = 0.40$ (Table III.2). In an additional evaluation setting assuming RR = 0.80 (data not shown), the effect of K, σ_c and the advantage of GEE versus the other methods was confirmed but at lower power levels than the ones obtained when RR = 0.70, e.g. power of all methods only reached or surpassed 80% when ($K = 40$, $\sigma_c \leq 0.15$) while only GEE reached 80% at ($K = 20$, $\sigma_c = 0.05$).

Table III.1: Width of the CI (mean and c.v.) of the rate ratios (RR) obtained by 5 analytical methods for overdispersed counts of cluster randomized trials, differing in: total number of clusters (*K*), level of between-cluster variation (σ_c) and degrees of imbalance[†], in 1000 replicates for t-test - GLMM and 300 replicates for Bayes-HM

			Balanced					Slightly imbalanced					Highly imbalanced				
σ_c	*K*		t-test	GEE-Emp	GEE-MB	GLMM	Bayes-HM	t-test	GEE-Emp	GEE-MB	GLMM	Bayes-HM	t-test	GEE-Emp	GEE-MB	GLMM	Bayes-HM
0.05	**10**	**mean**	0.49	0.37	0.37	0.53	0.51	0.49	0.36	0.37	0.53	0.51	0.57	0.34	0.36	0.55	0.54
		c.v.	25.2	25.2	24.8	15.1	13.3	27.1	28.6	27.1	16.3	14.5	29.2	37.0	31.1	19.1	17.88
	20		0.32	0.28	0.28	0.33	0.34	0.33	0.28	0.28	0.34	0.34	0.37	0.28	0.28	0.34	0.35
			17.2	17.2	17.0	10.9	8.3	17.5	18.0	17.8	11.4	8.4	21.4	23.0	21.6	13.8	12.9
	40		0.22	0.20	0.20	0.22	0.23	0.22	0.20	0.21	0.22	0.23	0.25	0.20	0.20	0.22	0.23
			11.2	11.2	11.2	7.7	7.0	12.0	12.3	12.3	8.1	7.8	14.5	14.5	14.0	9.6	8.5
0.15	**10**		0.63	0.48	0.48	0.64	0.60	0.64	0.48	0.48	0.65	0.61	0.68	0.46	0.47	0.66	0.63
			24.5	24.5	24.5	21.5	20.7	25.5	26.2	25.9	22.6	24.0	26.8	33.3	31.0	25.1	25.9
	20		0.41	0.37	0.37	0.42	0.41	0.42	0.36	0.36	0.41	0.40	0.46	0.37	0.37	0.43	0.42
			17.9	18.1	18.0	16.8	18.2	17.1	17.2	17.2	16.2	16.2	18.7	19.9	20.2	17.9	16.5
	40		0.28	0.27	0.27	0.28	0.28	0.29	0.27	0.27	0.28	0.28	0.31	0.27	0.27	0.29	0.28
			12.2	12.3	12.2	11.7	13.9	11.7	12.0	12.2	11.5	13.5	13.5	13.5	14.3	12.8	14.1
0.4	**10**		1.19	0.90	0.90	1.21	1.19	1.21	0.92	0.92	1.23	1.20	1.24	0.92	0.93	1.25	1.25
			26.9	26.9	26.9	25.4	32.6	26.4	26.6	27.1	25.1	32.3	26.5	27.8	31.7	25.6	30.7
	20		0.79	0.70	0.70	0.79	0.77	0.82	0.72	0.73	0.81	0.82	0.82	0.71	0.71	0.80	0.79
			19.6	19.6	19.6	17.4	19.7	19.0	19.1	20.1	16.6	19.2	18.7	19.2	23.3	16.9	21.7
	40		0.56	0.53	0.53	0.55	0.55	0.56	0.53	0.53	0.55	0.55	0.57	0.54	0.54	0.56	0.54
			14.1	14.1	14.1	11.5	15.2	13.9	14.0	14.9	11.3	14.8	14.5	14.6	17.0	12.0	16.1

[†] Imbalance around a mean cluster size of 30 individuals per cluster

Table III.2: Type I error rates (for relative rate of RR = 1) and statistical power (for RR = 0.7) of 5 analytical methods for overdispersed counts of cluster randomized trials, differing in: total number of clusters (*K*), level of between-cluster variation (σ_c) and degrees of imbalance[†],in:1000 replicates for t-test - GLMM and 300 replicates for Bayes-HM

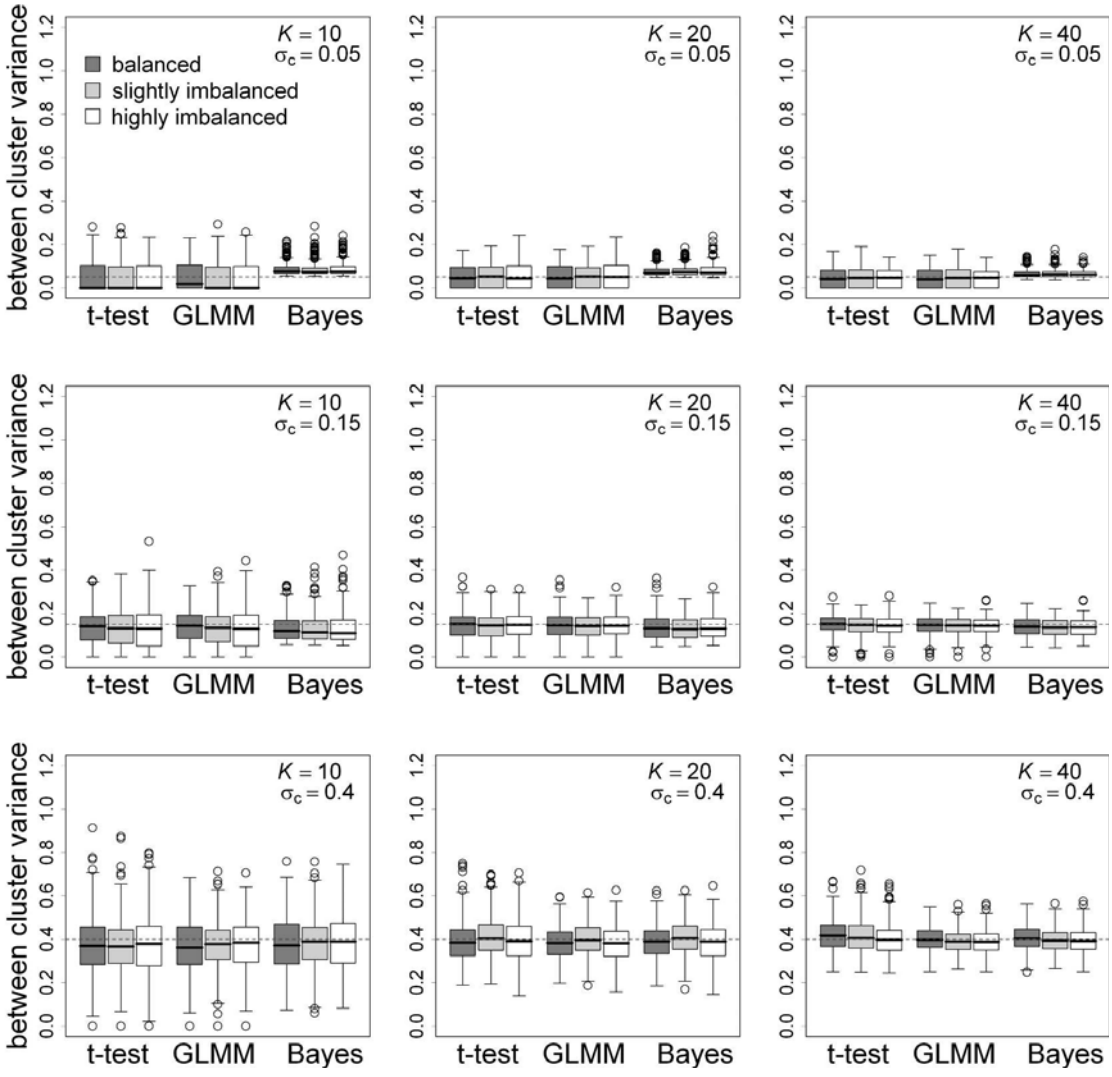
		Balanced					Slightly imbalanced					Highly imbalanced				
σ_c	*K*	t-test	GEE-Emp	GEE-MB	GLMM	Bayes-HM[‡]	t-test	GEE-Emp	GEE-MB	GLMM	Bayes-HM	t-test	GEE-Emp	GEE-MB	GLMM	Bayes-HM
Type I error rates																
0.05	**10**	0.046	**0.108**[*]	**0.102**	0.014	0.033	0.046	**0.144**	**0.140**	0.018	0.020	0.045	**0.203**	**0.168**	0.016	0.013
	20	0.058	**0.082**	**0.082**	0.030	0.037	0.042	**0.074**	**0.078**	0.030	0.033	0.050	**0.109**	**0.109**	0.027	0.027
	40	0.051	0.073	0.073	0.047	0.053	0.062	0.062	0.062	0.049	0.047	0.042	0.059	0.061	0.029	0.030
0.15	**10**	0.052	**0.118**	**0.116**	0.042	0.047	0.056	**0.128**	**0.132**	0.040	0.070	0.040	**0.150**	**0.138**	0.045	0.053
	20	0.066	**0.086**	**0.088**	0.068	0.060	0.058	**0.090**	**0.090**	0.054	0.070	0.041	**0.092**	**0.090**	0.048	0.047
	40	0.040	0.051	0.051	0.044	0.050	0.044	0.064	0.062	0.053	0.060	0.046	0.059	0.059	0.043	0.060
0.4	**10**	0.042	**0.146**	**0.148**	0.050	0.053	0.046	**0.140**	**0.138**	0.050	0.053	0.040	**0.132**	**0.136**	0.040	0.063
	20	0.054	**0.098**	**0.098**	0.064	**0.086**	0.054	**0.090**	**0.090**	0.050	0.067	0.041	**0.078**	**0.087**	0.045	0.067
	40	0.020	0.044	0.044	0.018	0.040	0.033	0.047	0.053	0.047	0.070	0.049	**0.067**	**0.074**	0.049	0.060
Statistical Power																
0.05	**10**	0.804	0.929	0.925	0.778	0.803	0.783	0.923	0.922	0.776	0.780	0.706	0.921	0.916	0.774	0.787
	20	0.991	0.995	0.995	0.991	0.993	0.988	0.998	0.998	0.992	0.997	0.939	0.996	0.995	0.989	0.990
	40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.15	**10**	0.588	0.789	0.788	0.584	0.663	0.580	0.764	0.768	0.585	0.590	0.512	0.784	0.771	0.575	0.590
	20	0.927	0.960	0.963	0.932	0.957	0.908	0.944	0.941	0.917	0.927	0.864	0.946	0.947	0.908	0.920
	40	0.998	0.999	0.999	0.999	0.990	1.000	1.000	1.000	1.000	1.000	0.993	1.000	1.000	0.999	1.000
0.4	**10**	0.196	0.403	0.398	0.214	0.217	0.202	0.401	0.405	0.224	0.227	0.181	0.359	0.367	0.201	0.223
	20	0.386	0.488	0.488	0.407	0.440	0.383	0.497	0.496	0.412	0.413	0.367	0.502	0.503	0.412	0.473
	40	0.697	0.737	0.742	0.739	0.677	0.664	0.712	0.720	0.701	0.657	0.667	0.716	0.716	0.696	0.683

[†]Imbalance around a mean cluster size of 30 individuals per cluster . [‡] Based on a Bayesian pseudo p-value computed as: 2*min[P(β > 0 | data); P(β < 0 | data)]. [*] Bold font indicates lower limit > than 0.05

Estimation of the between-cluster variance

The distribution of CV_c estimates from the t-test (ANOVA variance estimator in (10)), GLMM and Bayes-HM across replicate datasets is shown in Figure III.3; all negative variances were truncated to 0 for the t-test, and GLMM. Although the methods estimate on average the underlying σ_c, Bayes-HM produced in general more efficient estimates, becoming similar to GLMM but still superior to t-test when $\sigma_c = 0.4$. Note that between-cluster variance estimates for all the methods are greatly affected by K and σ_c yielding rather variable estimates at low K and high σ_c. Although with more homogeneous estimates, Bayes-HM overestimated the intercluster variance when $\sigma_c = 0.05$.

Figure III.3: Between-cluster coefficient of variation (CV_c) obtained by 3 analytical methods for cluster randomized trials, applied to 300 simulated datasets per combination of total number of clusters (K), between-cluster variation (σ_c) and degree of cluster size imbalance.



Analysis of the motivating example

The BoliviaWET trial initially targeted 30 children per cluster in 22 communities giving a total of 660 participants [35]. By the end of the trial however the observed cluster size was not constant but showed a symmetric distribution around a mean of 33 children per cluster (min = 23, median = 30, max = 57, s.d. = 7.6) and a total of 725 children recruited up to the randomization time: 349 and 376 children in the control and intervention arm, respectively.

A total of 887 diarrhoeal episodes were observed during 75,077 children-days at risk observed in the control arm and 808 diarrhoeal episodes in 82,682 children-days at risk observed in the intervention arm. It yields a crude RR of 0.827 corresponding to an effectiveness of 17.3% in reducing diarrhoea. The significance of the intervention effect was analysed by each of the 5 methods for clustered data examined above. 45,000 iterations after 2,000 burn-in were applied for the Bayes-HM.

In order to investigate the effects of overdispersion, two aspects were assessed:

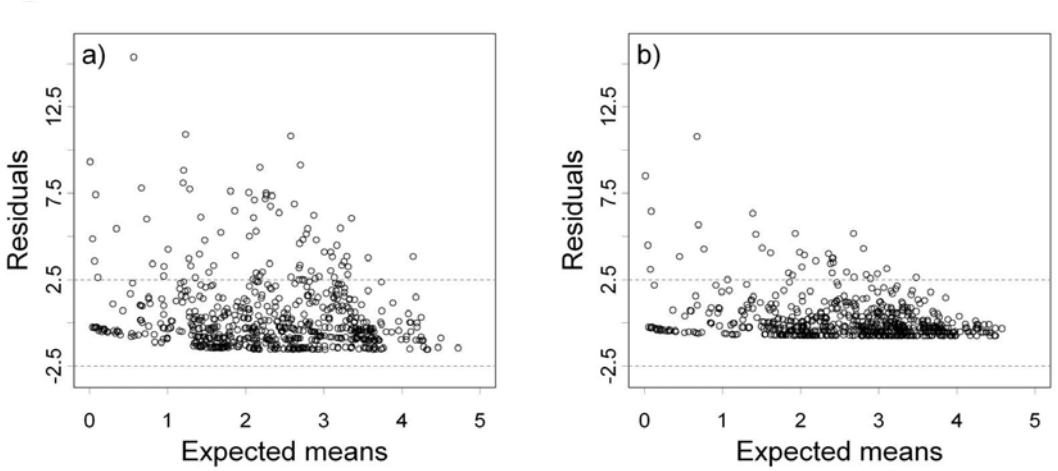
i) Both Poisson and NB mean and variance functions were specified (applicable on GEE, GLMM and Bayes-HM)
ii) Data were analyzed with or without the exclusion of outlier observations. Outliers were defined as those having the (PA) Pearson's standardized residuals greater than |2.5| for the model with the best fit. This left a remainder of 691 children (Table III.3).

The overdispersion parameter ϕ was always greater than 1 when Poisson variation was assumed even with the exclusion of outliers, while it draws close to 1 when NB variation was assumed (Table III.3), indicating that the NB model provides a better representation of the sampling variation. The distribution of residuals comparing the Poisson versus NB model confirms this result (Figure III.4).

The between-cluster coefficient of variation CV_c is presented for the t-test, estimates of σ_c for GLMM and Bayes-HM and within-cluster exchangeable correlation α for

GEE (Table III.3). A cluster variance (correlation for GEE) higher than 0 was obtained by all the methods, under either Poisson or NB distributions and with or without the exclusion of outliers. The only exception was the t-test in the complete dataset (N = 725; Table III.3). This null clustering is explained by the negative variance (truncated to 0) estimated as $(MS_c - MS_e)/n_0$ (see expression (10)), where a high residual variability, captured by the MS_e exceeded the clusters'. Indeed, when the outliers were excluded (N = 691) the method estimated an 18.8% of between-cluster variation with MS_c becoming clearly higher than MS_e. That suggests that using this method the outliers contribute more to the residual than to the cluster variance. Note that during the simulations the estimate of the between-cluster variance was particularly unstable when $\sigma_c > 0.15$.

Figure III.4: a) Poisson and b) Negative Binomial (Pearson-standardized) residuals versus the expected mean number of events of the BoliviaWET dataset.



The cluster variation estimated by adjusting the standard error by ϕ in the Poisson models (GLMM2 in Table III.3) decreased when comparing the uncorrected with the corrected Poisson GLMM. A portion of the cluster variance of the uncorrected model went thus to adjust the standard errors in the GLMM2. Finally, the posterior medians of σ_c of the Bayes-HM are similar to the values estimated by the equivalent GLMM model.

Table III.3: Parameter estimates of the analysis of the BoliviaWET trial, obtained by 5 statistical methods, with and without the exclusion of outliers, and assuming Poisson or Negative Binomial distributed counts.

Parameter	Complete dataset, N = 725									
		Poisson					Negative Binomial			
	t-test	GEE-Emp	GEE-MB	GLMM1	GLMM2†	Bayes-HM	GEE-Emp	GEE-MB	GLMM1	Bayes-HM
ϕ	-	5.95		4.74	4.95	-	1.53		1.28	-
NB parm (s)	-	-		-	-	-	1.33		1.42	1.26
$CV_c \mid \alpha \mid \sigma_c$	0.000	0.037		0.324	0.195	0.330	0.016		0.270	0.271
RR	0.908	0.921	0.921	0.868	0.853	0.865	0.843	0.843	0.809	0.806
CI of RR	(0.61, 1.20)	(0.70, 1.21)	(0.64, 1.32)	(0.64, 1.18)	(0.64, 1.14)	(0.63, 1.18)	(0.64, 1.11)	(0.63, 1.13)	(0.59, 1.12)	(0.59, 1.10)
Width of the CI	0.59	0.50	0.67	0.54	0.50	0.55	0.47	0.51	0.53	0.52
p-value‡	0.496	0.550	0.653	0.346	0.262	0.343	0.225	0.259	0.187	0.172

Parameter	Outliers excluded*, N = 691									
		Poisson					Negative Binomial			
	t-test	GEE-Emp	GEE-MB	GLMM1	GLMM2	Bayes-HM	GEE-Emp	GEE-MB	GLMM1	Bayes-HM
ϕ	-	2.87		2.62	2.66	-	0.97		1.02	-
NB parm (s)	-	-		-	-	-	0.93		0.79	0.90
$CV_c \mid \alpha \mid \sigma_c$	0.188	0.022		0.262	0.195	0.265	0.019		0.195	0.163
RR	0.924	0.885	0.885	0.906	0.890	0.902	0.887	0.887	0.887	0.876
CI of RR	(0.66, 1.19)	(0.70, 1.12)	(0.70, 1.12)	(0.64, 1.18)	(0.64, 1.14)	(0.63, 1.18)	(0.71, 1.11)	(0.71, 1.11)	(0.69, 1.14)	(0.69, 1.11)
Width of the CI	0.53	0.42	0.42	0.48	0.46	0.48	0.41	0.40	0.46	0.42
p-value	0.539	0.307	0.311	0.437	0.352	0.414	0.301	0.293	0.338	0.281

ϕ = Overdispersion parameter, estimated as the generalized Pearson chi-square statistics

NB parm (s) = Scale (overdispersion) parameter of the Negative Binomial distribution

Between-cluster coefficient of variation CV_c is reported for the t-test, α the exchangeable correlation for GEE and σ_c for GLMM and Bayes-HM

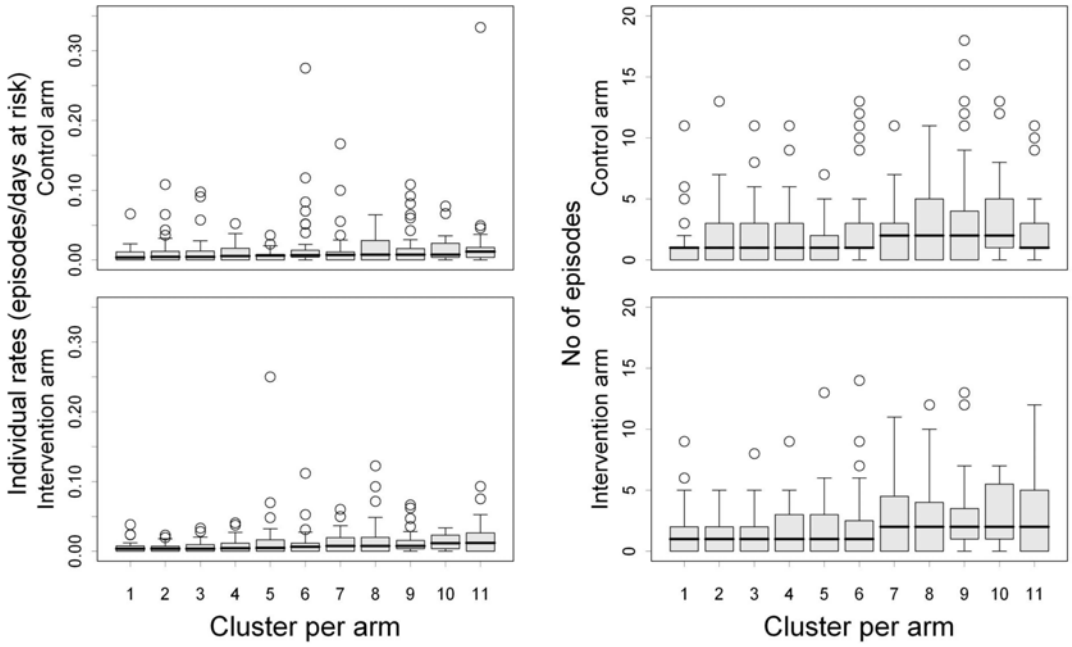
† GLMM2= GLMM standard errors corrected (inflated) by ϕ

* Excluding observations whose absolute standardized residuals were higher than |2.5|

‡ Bayesian pseudo p-value computed as: 2*min[P(β > 0 | data); P(β < 0 | data)]

Although the simulations gave similar results for different estimation methods, the assumed outcome distribution had an important influence on the point estimate of the RR in the complete BoliviaWET dataset. All the Poisson-based methods and the t-test gave RR above 0.85 with the overdispersion-corrected Poisson GLMM showing the closest RR to the crude 0.827. The NB models yielded estimates lower than the Poisson models, more homogeneous and much closer to the observed (population-averaged) crude RR. When outliers were excluded from the analysis the RR of all the methods moved the RR estimates towards unity. As shown in Figure III.5, this behaviour may be explained by the fact that the control arm had most of the observed outliers, particularly in the Poisson model. Their exclusion reduced the estimated difference between control and intervention rates.

Figure III.5: Distribution of the individual incidence rates and the number of episodes per community in the two study arms of the BoliviaWET trial.



As expected, the CI's were narrower when outliers were excluded as a result of the reduction in the overall variance. In general, the exclusion of outliers leads to more similarities with the simulation results, both for Poisson and NB analysis, with the GEE methods showing narrower CI than the other methods. This behaviour however disappears in the complete dataset, where substantial differences between the CI widths between the GEE-Emp and GEE-MB were observed. Further, the t-test CI's

widths were very different from those of GLMM and Bayes-HM analysis. The latter result may be because of the highly variable widths of the t-test CIs compared with those of GLMM and Bayes-HM analyses, as observed in the simulations (Table III.1).

No significant effect of the intervention on the diarrhoeal rates was found by any method in any scenario. The interval estimates all contained the null effect and the P-values were all above 0.15, although a tendency towards even lower significance was found for all the methods when N = 691, since the excluded outliers belonged mainly to the control arm (Figure III.5). The NB models generated less conservative results for the complete data set, but seemed to best model the RRs even though outliers were present. Note that the (overdispersion-adjusted) Poisson GLMM2 applied to N = 725 produced close results to the NB in terms of estimates and significance.

According to the simulations, GEE has more power than the other methods to detect true significant results. Any of the methods would have ≥ 80% power to detect the 33% difference initially planned in the BoliviaWET trial with 22 clusters and a σ_c of 0.27 (assuming the NB model on the complete dataset) (see Table III.2), but not for the observed 17.7% crude rate. The simulations with RR = 0.80 (results not shown) suggest that GEE with NB functions would have ≈ 70% power while GLMM ≈ 60% to detect a 20% reduction. The Bayesian posterior probabilities $Pr[\exp(\beta)>d \mid \text{data}]$ (with d as the effect of interest) concur with this estimates yielding powers of 0.82 and 0.52 for d=0.7 and d=0.8.

5. DISCUSSION

The performance of analytical methods for overdispersed count data in cluster randomized trials was examined in terms of point, interval estimation and hypothesis testing. The methods were: the two-sample t-test of cluster-level incidence rates, GEE with empirical covariance estimators, GEE with model-based covariance estimators, GLMM and Bayesian hierarchical models under negative binomial distribution when applicable. We focused on overdispersed counts, allowing for variation in times of individual follow-up and simulated situations close to reality for community-

intervention trials, considering the effects of: small number of clusters (*K*), different degrees of clustering (σ_c) and different levels of cluster size variation. Overdispersion was stressed through the data analysis of a community-intervention trial to illustrate its impact in the performance of the statistical methods.

The performance of the methods was related to *K* and σ_c while high imbalance affected the performance of the t-test and somewhat reduced the already low coverage of GEE in small samples.

Imbalance and cluster size influence the performance of methods for binary clustered data [22, 31, 32, 50-52], particularly of cluster-level methods (e.g. t-test, cluster-level linear regretssion), but do not affect individual level random-effect models for binary data [53] . We found similar results when analysing count data. One particular study [31], assuming conditions typical to primary care trials, shows that power is affected when the cluster size coefficient of variation (c.v.) is greater than 0.23. Our findings assuming community field trials situations are in line with that conclusion. Note however that the mechanisms that determine variable cluster size in community randomized trials may differ from their primary care counterparts. In community trials, the investigators often have more control over the size of the clusters, because there is generally a choice in how to subdivide the population into communities (e.g. geographic areas, villages, districts) [3]. The underlying distribution of community size and the patterns of individuals' response/drop-outs are important sources of cluster size variation in field trials. On the contrary in primary care, the recruitment strategy of individuals or clusters may be more important (e.g. health care seeking, degree of disease register size), leading thus to greater variation in the cluster size. We chose the slightly imbalance scenario (c.v. = 0.2) to match the BoliviaWET experience of failing to recruit equally sized clusters. We believe this is common in community randomized trials. The high imbalance (c.v.=0.6) represents situations where half the clusters have sizes lower or larger than the minimum and maximum cluster size in the BoliviaWET trial. This choice matches the average level of imbalance of health facilities in the UK [31]. We interpret our findings as applicable to trials with a mean cluster size of 30, which appears to be the average size of

community-intervention trials < 100 individuals/cluster (confirmed by an *adhoc* review of 20 community-intervention trials published after 2000).

All the methods are similar in terms of point estimation. Theoretically no important differences were to be expected particularly between PA and CS models, since for count data under a log link, CS's regression coefficient have both PA and CS interpretations [23, 27]. Indeed, the observed differences in bias and EmpSE depended only on K and σ_c.

Although we present results for the 30% but analysed also the 20% reduction in the true RR, we found GEE methods to have higher power than the other methods, in line with previous research [18, 21, 54]. However, this advantage was clear for $K < 40$, the same region where the CI's coverage was anticonservative, and in some extent at $K = 40$ when $\sigma_c = 0.4$. In the analysis of the 20% reduction in the RR, GEE's higher power was nonetheless confirmed for $K = 40$ when $\sigma_c > 0.05$.

GEE-Emp and GEE-MB produced noticeable lower coverage probabilities alongside narrow CI and consequently higher Type I error rates compared to the other methods. This findings are consistent with previous research [17, 18, 22, 50, 54, 55] and may partly be explained by the fact that GEE intervals are based on normal quantiles, while t-test and GLMM base their CI on the student's distribution, more appropriate for small sample situations. In addition, GEE have been reported to underestimate the covariance among observations producing downwards biased standard errors in small sample situations (< 40 clusters), specially with unequal cluster sizes [56]. Bias-corrected methods have accordingly been proposed [56-58], although they are not yet implemented in standard statistical packages. Simple sampling distribution corrections are however possible and have been shown to improve GEE performance [22, 59]. An advantage of GEE-MB over GEE-Emp in small samples is expected if the correlation structure is correctly specified. We could not confirm this because in terms of correlation structure and under the log link, the CS underlying model used to generate the data is not equivalent to the exchangeable PA [27]. We applied the exchangeable nonetheless because of its common use in CRT. Recent research proposes the means to identify the working correlation structure [60].

Although Bayes-HM reported higher Type I error rates than GLMM when σ_c was high both methods performed similarly well in all the studied scenarios. Proper coverage performance in simulations of clustered data has been described for both methods with some advantages of Bayes-HM over GLMM [50]. The similarities in interval widths between GLMM and Bayes-HM that we found do not concur with Turner et al's claim [13] that Bayesian hierarchical models produce wider CI for β than frequentist multilevel models in an analysis of binary outcomes, since the Bayesian models account for imprecision of the intercluster variance while the frequentist models assume it to be known. In the present study, the analysis of count data in WinBugs using the same priors as [13] and applied to 300 different datasets per arrangement produced consistently similar intervals to those obtained by the GLMM in the GLIMMIX procedure. Although the methods are similar, Bayes-HM has the advantage of greater flexibility in assessing diverse outcome and random-effect distributions, and provides interval estimates for any parameter or function of parameters of interest such as the intraclass correlation coefficient (ICC)) or the statistical power. However informative priors should only be used with caution because of their influence on the uncertainty measures [13, 42].

The straightforward t-test showed conservative results, wide CI, rather large coverage probabilities, and a tendency to have lower power than GLMM and Bayes-HM at the highest between-cluster variability (σ_c=0.40). Its disadvantages were evident in the analysis of the BoliviaWET dataset where it reduced the difference between study-arms by masking the effect of outlier rates which mainly came from the control arm (Figure III.5). Another limitation of all cluster-level methods regards the inability of adjusting for individual-level covariates.

We investigated the estimates of the random-effects, given their implications for the ICC, CV_c and the design of new trials, but did not compare the ICC from the random-effects models with the exchangeable correlations of the GEE because of the underlying differences mentioned above. The relationships between CS and PA in terms of marginal covariances and correlations for count data have been presented elsewhere for the case of Poisson variance [27]. A derivation of similar equivalences

would be required for Negative Binomial but that is out of the scope of this study. Random-effects models have also previously been compared with GEE in terms of variance and covariance parameter estimates in continuous and binary data [61]. In general the t-test, GLMM and Bayes-HM all provided reasonably good estimates of the intercluster coefficient of variation CV_c for $K \geq 20$ but all methods gave a high dispersion when the true value of $\sigma_c = 0.4$ or $K = 10$. Bayes-HM produced in general more stable values but upward-biased when $\sigma_c = 0.05$. The classical ANOVA estimator was the more unstable and may produce misleading results in presence of extreme observations as was observed in the motivating dataset.

The analysis of the BoliviaWET dataset illustrates the impact of modelling extra-Poisson variation: a situation that routinely occurs in count data of CRTs. Overdispersion may make itself evident both as inflation of the incidence of zero counts or occurrence of larger counts than expected by the Poisson model and is known to cause underestimation of standard errors and misleading inference for the regression parameters [62]. At the same time, it is important to distinguish real from apparent overdispersion that can arise *inter alia*, because of omission of explanatory predictors and/or interactions, presence of outliers, or miss-specification of the link [38]. However, some proposed remedies for apparent overdispersion are not applicable to CRTs. For instance, no other predictors than the design and treatment factors are included in the analysis of a crude model, and dropping/adjusting for outliers would infringe the principle of the intention to treat analysis.

Approaches to deal with real overdispersion such as inflating the Poisson variance by ϕ or assuming a heterogeneous gamma-distributed Poisson mean [63] (Poisson-Gamma mixture [38]) may not be enough for a CRT. Including cluster random-effects in a Poisson CS model implies that overdispersion is assumed [27], but this approach does not necessarily capture the individual within-cluster heterogeneity. In the BoliviaWET data, despite some extreme observations, the number of outliers in the NB analysis was clearly lower than that of the Poisson model (Figure III.4), in addition, NB showed a superior fit even without adding predictors other than the treatment. The NB model with normally distributed random-effects would be thus preferred to address overdispersion in a CRT. It is comparable to the Poisson model

with gamma and normal random-effects at mean and cluster levels respectively, which has previously been shown to give improved fit over Poisson, Poisson-gamma and Poisson-normal models when estimated via full maximum likelihood with numerical integration over the random-effects [26] (PROC NLMIXED in SAS). In contrast, we used expansion methods (PROC GLIMMIX) and MCMC (Bayes-HM) which have the advantage over PROC NLMIXED because they can be extended to pair-matched, repeated cross-sectional or other more complex designs.

To our knowledge this is the first study that has used simulation to evaluate analytical methods for overdispersed counts in CRTs. There is still a need to consider more complex designs (pair-matching, stratified, repeated cross-sectional), and to asses imbalance under different average cluster sizes. Further research is needed into exploring the implications of different degrees of overdispersion. We did not evaluate other extra-Poisson models (Zero Inflated, Zero Truncated models for Poisson and NB Regression) and did not analyse the effects of covariate inclusion in the context of borderline overdispersion in the adjusted analysis of CRTs. Another important limitation is the use of only 300 datasets per arrangement for Bayes-HM because of the long computation times. Statistical power was reported only for one treatment difference (30%) although results were confirmed with a lower treatment difference (20%).

Under the situation of community-intervention trials analysed in this paper, our overall conclusions are that the NB model with normal random-effects provides a natural way to address overdispersion of count data in a CRT. Its analysis via GLMM and Bayes-HM would produce overall good performance, although caution must be taken for the random-effects estimates when $K = 10$ or $\sigma_c = 0.4$. GEE with NB means and variance functions are also an attractive choice provided its higher power. GEE requires however a proper specification of the correlation structure in small-sample situations, which in practice may differ from the structures assumed by the typically employed exchangeable and/or use of bias-corrected estimators. Based on our simulations the t-test is conservative for overdispersed rates and caution must be taken when extreme observations are present. High imbalance affects the overall performance of the t-test cluster-level analysis and coverage of GEE when $K = 10$.

Acknowledgments

We are grateful to two anonymous reviewers for their valuable insight that greatly improved this paper. We are also thankful to Dr. Laura Gosoniu and Benjamin Arnold for their comments on a previous version of the manuscript. Finally we warmly acknowledge Andri Christen, Myriam Cevallos, Freddy Arauco and the rest of the field staff of the BoliviaWET study for their relentless commitment during the data collection. The BoliviaWET trial was funded by the National Institutes of Health, award number R01AI50087-01IH. Gonzalo Durán Pacheco is in receipt of a stipend from the Stipendiumkommission of the Amt für Ausbildungsbeiträge of the Canton of Basel, Switzerland.

6. REFERENCES

1. Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin.Trials* 2004; **1**(1):80-90.

2. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: A review of recent methodological developments. *American Journal of Public Health* 2004; **94**(3):423-432.

3. Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials and Statistics in Medicine. *Statistics in Medicine* 2007; **26**(1):2-19.

4. Donner A, Klar N. Methods for Comparing Event Rates in Intervention Studies When the Unit of Allocation Is A Cluster. *American Journal of Epidemiology* 1994; **140**(3):279-289.

5. Klar N, Darlington G. Methods for modelling change in cluster randomization trials. *Statistics in Medicine* 2004; **23**(15):2341-2357.

6. Bennett S, Parpia T, Hayes R, Cousens S. Methods for the analysis of incidence rates in cluster randomized trials. *International Journal of Epidemiology* 2002; **31**(4):839-846.

7. Preisser JS, Young ML, Zaccaro DJ, Wolfson M. An integrated population-averaged approach to the design, analysis and sample size determination of cluster-unit trials. *Statistics in Medicine* 2003; **22**(8):1235-1254.

8. Twisk JWR. *Applied Multilevel Analysis: A Practical Guide for Medical Researchers*. Cambridge University Press: UK, 2006.

9. Greenland S. When should epidemiologic regressions use random coefficients? *Biometrics* 2000; **56**(3):915-921.

10. McCulloch CHE, Searle SR. *Generalized, Linear, and Mixed Models*. Wiley & Sons: New York, 2001.

11. Zeger SL, Liang KY, Albert PS. Models for Longitudinal Data - A Generalized Estimating Equation Approach. *Biometrics* 1988; **44**(4):1049-1060.

12. Spiegelhalter DJ. Incorporating Bayesian ideas into health-care evaluation. *Statistical Science* 2004; **19**(1):156-174.

13. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* 2001; **20**(3):453-472.

14. Ohlssen DI, Sharples LD, Spiegelhalter DJ. Flexible random-effects models using Bayesian semi-parametric models: Applications to institutional comparisons. *Statistics in Medicine* 2007; **26**(9):2088-2112.

15. Neuhaus JM, Kalbfleisch JD, Hauck WW. A Comparison of Cluster-Specific and Population-Averaged Approaches for Analyzing Correlated Binary Data. *International Statistical Review* 1991; **59**(1):25-35.

16. Albert PS, McShane LM. A generalized estimating equations approach for spatially correlated binary data: applications to the analysis of neuroimaging data. *Biometrics* 1995; **51**(2):627-638.

17. Hendricks SA, Wassell JT, Collins JW, Sedlak SL. Power determination for geographically clustered data using generalized estimating equations. *Statistics in Medicine* 1996; **15**(17-18):1951-1960.

18. Bellamy SL, Gibberd R, Hancock L, Howley P, Kennedy B, Klar N, Lipsitz S, Ryan L. Analysis of dichotomous outcome data for community intervention studies. *Statistical Methods in Medical Research* 2000; **9**(2):135-159.

19. Pan W, Wall MM. Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. *Stat.Med.* 2002; **21**(10):1429-1441.

20. Yasui Y, Feng ZD, Diehr P, McLerran D, Beresford SAA, McCulloch CE. Evaluation of community-intervent ion trials via generalized linear mixed models. *Biometrics* 2004; **60**(4):1043-1052.

21. Austin PC. A comparison of the statistical power of different methods for the analysis of cluster randomization trials with binary outcomes. *Statistics in Medicine* 2007; **26**(19):3550-3565.

22. Ukoumunne OC, Carlin JB, Gulliford MC. A simulation study of odds ratio estimation for binary outcomes from cluster randomized trials. *Statistics in Medicine* 2007; **26**(18):3415-3428.

23. Ritz J, Spiegelman D. Equivalence of conditional and marginal regression models for clustered and longitudinal data. *Statistical Methods in Medical Research* 2004; **13**(4):309-323.

24. Chin HC, Quddus MA. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis and Prevention* 2003; **35**(2):253-259.

25. Tseloni A. Multilevel modelling of the number of property crimes: household and area effects. *Journal of the Royal Statistical Society Series A-Statistics in Society* 2006; **169**:205-233.

26. Molenberghs G, Verbeke G, Demetrio CGB. An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis* 2007; **13**(4):513-531.

27. Young ML, Preisser JS, Qaqish BF, Wolfson M. Comparison of subject-specific and population averaged models for count data from cluster-unit intervention trials. *Statistical Methods in Medical Research* 2007; **16**(2):167-184.

28. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? *Stat.Med.* 2009; **28**(2):221-239.

29. Kerry SM, Bland JM. Unequal cluster sizes for trials in English and Welsh general practice: implications for sample size calculations. *Stat.Med.* 2001; **20**(3):377-390.

30. Yudkin PL, Moher M. Putting theory into practice: a cluster randomized trial with a small number of clusters. *Statistics in Medicine* 2001; **20**(3):341-349.

31. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International Journal of Epidemiology* 2006; **35**(5):1292-1300.

32. van Breukelen GJP, Candel MJJM, Berger MPF. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Statistics in Medicine* 2007; **26**(13):2589-2603.

33. Sommer B, Marino A, Solarte Y, Salas ML, Dierolf C, Valiente C, Mora D, Rechsteiner R, Setter P, Wirojanagud W, Ajarmeh H, AlHassan A, Wegelin M. SODIS - An emerging water treatment process. *Journal of Water Supply Research and Technology-Aqua* 1997; **46**(3):127-137.

34. Dejung S, Fuentes I, Almanza G, Jarro R, Navarro L, Arias G, Urquieta E, Torrico A, Fenandez W, Iriarte M, Birrer C, Stahel WA, Wegelin M. Effect of solar water disinfection (SODIS) on model microorganisms under improved and field SODIS conditions. *Journal of Water Supply Research and Technology-Aqua* 2007; **56**(4):245-256.

35. Mäusezahl, D., Christen, A., Duran-Pacheco, G., Alvarez-Tellez, F., Iriarte, M., Zapata M.E., Cevallos, M., , Hattendorf J., M., Arnold, B., Smith-A T.,

and Colford, J. M. A cluster-randomized, controlled trial of solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia. 2009 (submitted manuscript) .

36. Heeren T, D'Agostino R. Robustness of the two independent samples t-test when applied to ordinal scaled data. *Stat.Med.* 1987; **6**(1):79-90.

37. McCullagh P, Nelder JA. *Generalized Linear Models*. Chapman and Hall: London, 1989.

38. Hilbe J.H. *Negative Binomial Regression*. Cambridge University Press, New York: 2007.

39. Rodriguez G, Goldman N. An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. *Journal of the Royal Statistical Society Series A-Statistics in Society* 1995; **158**:73-89.

40. Pinheiro JC, Bates DM. Approximations to the Log-likelihood Function in the Nonlinear Mixed-effects Model. *Journal of Computational andGraphical Statistics* 1995; **4**:12-35.

41. SAS Institute Inc. *The GLIMMIX Procedure*. SAS Institute Inc.: Cary, North Carolina, USA, 2006.

42. Spiegelhalter DJ. Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* 2001; **20**(3):435-452.

43. Young M.L. *Generalized estimating equations (GEE) with design-based correlation structures for cluster-unit trials*. University of North Carolina: 2003.

44. Liang KY, Zeger SL. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; **42**(1):121-130.

45. Fitzmaurice G. M, LairdN.M., Ware J.H. *Applied longitudinal analysis*. Wiley: 2004.

46. Horton NJ, Lipsitz SR. Review of software to fit generalized estimating equation regression models. *American Statistician* 1999; **53**(2):160-169.

47. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *International Journal of Epidemiology* 1999; **28**(2):319-326.

48. SAS Institute Inc. *SAS/STAT 9.1 user's guide*. SAS institute Inc.: Cary: NC, 2004.

49. Ukoumunne OC. A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. *Stat.Med.* 2002; **21**(24):3757-3774.

50. Localio AR, Berlin JA, Ten Have TR. Longitudinal and repeated cross-sectional cluster-randomization designs using mixed effects regression for

binary outcomes: Bias and coverage of frequentist and Bayesian methods. *Statistics in Medicine* 2006; **25**(16):2720-2736.

51. Klar N, Donner A. Current and future challenges in the design and analysis of cluster randomization trials. *Statistics in Medicine* 2001; **20**(24):3729-3740.

52. Guittet L, Ravaud P, Giraudeau B. Planning a cluster randomized trial with unequal cluster sizes: practical issues involving continuous outcomes. *BMC.Med.Res.Methodol.* 2006; **6**:17.

53. Heo M, Leon AC. Performance of a mixed effects logistic regression model for binary outcomes with unequal cluster size. *Journal of Biopharmaceutical Statistics* 2005; **15**(3):513-526.

54. Heo M, Leon AC. Comparison of statistical methods for analysis of clustered binary observations. *Statistics in Medicine* 2005; **24**(6):911-923.

55. Gunsolley JC, Getchell C, Chinchilli VM. Small Sample Characteristics of Generalized Estimating Equations. *Communications in Statistics-Simulation and Computation* 1995; **24**(4):869-878.

56. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001; **57**(1):126-134.

57. Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 2001; **96**(456):1387-1396.

58. Lu B, Preisser JS, Qaqish BF, Suchindran C, Bangdiwala S, Wolfson M. A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics* 2007; **63**(3):935-941.

59. Lipsitz SR, Fitzmaurice GM, Orav EJ, Laird NM. Performance of Generalized Estimating Equations in Practical Situations. *Biometrics* 1994; **50**(1):270-278.

60. Hin LY, Wang YG. Working-correlation-structure identification in generalized estimating equations. *Stat.Med.* 2009; **28**(4):642-658.

61. Evans BA, Feng ZD, Peterson AV. A comparison of generalized linear mixed model procedures with estimating equations for variance and covariance parameter estimation in longitudinal studies and group randomized trials. *Statistics in Medicine* 2001; **20**(22):3353-3373.

62. Hinde J, Demetrio CGB. Overdispersion: Models and estimation. *Computational Statistics & Data Analysis* 1998; **27**(2):151-170.

63. Lee Y, Nelder JA. Two ways of modelling overdispersion in non-normal data. Journal of the Royal Statistical Society Series C-Applied Statistics 2000; 49:591-598.

Chapter IV:

Point and Interval estimation of the between-cluster coefficient of variation for overdispersed counts in cluster randomized trials

Gonzalo Durán Pacheco[1], Jan Hattendorf[1], John M. Colford, Jr.[2], Daniel Mäusezahl[1], Thomas Smith[3]

[1] *Department of Public Health and Epidemiology, Interventions and Health Systems Unit, Swiss Tropical Institute*
[2] *Division of Epidemiology, School of Public Health, University of California, Berkeley (UCB), California*
[3] *Department of Public Health and Epidemiology, Biostatistics and Epidemiology Unit, Swiss Tropical Institute, P.O. box 4002 Basel, Switzerland.*

This manuscript is prepared to be submitted to: *Biometrics*

SUMMARY

We studied the estimation of the between-cluster coefficient of variation of overdispersed counts, as a measure to assess clustering in community randomized trials. Four methods for obtaining point estimates and three methods for interval estimation were assessed via simulation under different sample sizes and levels of clustering. The point estimating methods were: i) a cluster-level coefficient of variation (CL), ii) one-way random effects ANOVA, iii) generalized linear mixed models (GLMM) iv) Bayesian hierarchical models (Bayes-HM), the last two assuming Negative Binomial distribution. The interval estimating methods were: i) Bootstrap CI, ii) Generalized CI and iii) Bayes-HM. GLMM and ANOVA both provided unbiased point estimates although ANOVA was more unstable under high clustering. CL heavily overestimated the between-cluster variation when it is lower or equal to 25%. Bayes-HM provided slight upward bias in settings without clustering. Bayes-HM performed best in terms of interval estimation. We illustrate and discuss the application of these methods using data of a community randomized trial of solar water disinfection in rural Bolivia.

Keywords: Between-cluster coefficient of variation, confidence intervals, community-cluster randomized trials, Negative Binomial data.

1. INTRODUCTION

In randomized controlled trials of health interventions it is sometimes necessary to allocate interventions by groups (clusters) rather than at the individual level. Typical situations include *i)* interventions aimed at cluster level (communities, hospitals, general practices, schools, etc.), *ii)* behavioural change is desired at community level, *iii)* a need to avoid contamination in unblinded studies, *iv)* logistical convenience, among others. These trials are known as Cluster Randomized trials (CRTs) [1].

In CRTs, the similarity of individuals' responses within the same cluster invalidates the assumption of independence of standard statistical methods [2]. Sample size calculations and statistical analysis would therefore require adjustment for intracluster dependence to avoid an elevated type 2 error at the design stage or an inflated type 1 error at the analysis stage [3].

The common measures of clustering in CRTs are the *intraclass (intracluster) correlation coefficient* (ρ) and the *between-cluster coefficient of variation* (CV_c). Methods for point and interval estimation of ρ for continuous and binary data have been studied and critically reviewed [4-12]. Much less literature exists for CV_c, though it is generally easier to understand for field epidemiologists. Methods for determining sample size using CV_c as the measure of clustering have been described [13]. From the analytical point of view, when modelling count data by random-effects models, the square root of the cluster-effect variance approximately equals CV_c when the log link function is used. In terms of interval estimation, a number of studies present computationally cumbersome methods for confidence intervals (CI) of coefficient of variations, assuming normally distributed data [14, 15]. A much simpler approach based on the concept of generalized variables can be applied [16, 17].

In this paper we study the performance of methods for estimating CV_c for CRTs with overdispersed counts, motivated by the analysis of a community randomized trial of solar water disinfection in rural Bolivia. We compare: i) the coefficient of between-cluster variation of cluster-level rates [13], ii) the ANOVA variance component

estimator [5], iii) GLMM of Negative Binomial count data and the iv) variance component of Bayes-HM. We further assess interval estimation of CV_c linked to the methods above by applying i) Bootstrapping, ii) the CIs of generalized pivots and iii) Bayesian credible regions. Performance is assessed via Monte Carlo simulation with different sample sizes and degrees of clustering.

We introduce first the motivating example. Notation and the details of the methods applied throughout the paper are given in section 3 as well as a description of the simulation study. The findings are reported in section 4 together with the analysis of the example. Further connotations and conclusive remarks are commented in section 5.

2. MOTIVATING EXAMPLE

Solar drinking water disinfection (SODIS) is a low-cost, point-of-use water purification method that uses solar energy to inactivate waterborne pathogens. The method has been proven to be efficacious under lab conditions [18, 19], but evidence of its effectiveness in populations consuming contaminated water is scarce [20]. A community randomized trial (BoliviaWET) was conducted in 22 communities in rural Bolivia to evaluate the effect of a SODIS promotion campaign in reducing diarrhoea among children under 5 years of age [21]. Communities were pair-matched by baseline diarrhoeal incidence and the intervention was randomly allocated to one community within each pair. Diarrhoea was recorded daily by a surveillance monitoring system for one year. In this paper we analyze the effects of the intervention on the primary outcome expressed as the number of episodes per child (Y) per time at risk (t) and estimate the between-cluster variation ignoring pair-matching. We use the data of the trial for illustration purposes only.

3. METHODS

3.1. Estimating the between-cluster coefficient of variation CV_c

We denote Y_{ijl} the outcome of a CRT observed on individual l (l=1,…,n_{ij}), from cluster j (j=1,…,k_i), receiving the intervention i (i=1,2). Denoting the population cluster-level means by μ_j and the cluster variance $V(\mu_j)$, the between-cluster coefficient of variation is defined as:

$$CV_c = \frac{\sqrt{V(\mu_j)}}{\mu} \tag{1}$$

where $\mu = E(\mu_j)$. A common value of CV_c is assumed for both trial arms.

3.1.1. Estimating CV_c from the cluster-level rates.

Let us assume that the outcome variable takes values 0, 1, 2,…, with different periods of observation t_{ijl} among individuals. Defining the cluster-level rates as

$$r_{ij} = \sum_{l=1}^{n_{ij}} Y_{ijl} \Big/ \sum_{l=1}^{n_{ij}} t_{ijl} \ , \tag{2}$$

a first method of estimation [13] considers the cluster variance:

$${}^1\hat{V}\left(\mu_j\right) = s^2 - \frac{\bar{r}_{..}}{n} \sum_{i=1}^{2} \sum_{j=1}^{n_i} 1/r_{ij} \ , \tag{3}$$

where $s^2 = \frac{1}{K-1} \sum_{i=1}^{2} \sum_{j=1}^{k_i} \left(r_{ij} - \bar{r}_{..}\right)^2$, $\bar{r}_{..} = \frac{1}{K} \sum_{i=1}^{2} \sum_{j=1}^{k_i} r_{ij}$, $K = k_1 + k_2$. The coefficient of variation is given by:

$${}^1CV_c = \frac{\sqrt{{}^1\hat{V}\left(\mu_j\right)}}{\bar{r}_{..}} \tag{4}$$

3.1.2. The one way random-effects ANOVA estimator.

The one way random-effects model of the individual event rates $X_{jl} = Y_{jl}/t_{jl}$ is,

$$X_{jl} = \mu + r_j + \varepsilon_{jl} \tag{5}$$

with μ as the population mean, r_j the random effect of cluster $j \sim (0, \sigma_r^2)$ and ε_{jl} the random effect of individual l from cluster $j \sim (0, \sigma_e^2)$. r_j and ε_{jl} are usually assumed to be normally distributed, although this is not important for variance component point estimation [22]. When applied to event rates of overdispersed counts, this approach was reported to produce consistent point estimates of CV_c [23].

The between-cluster variance is estimated form the corresponding ANOVA table as:

$$\hat{\sigma}_r^2 = \frac{MS_c - MS_e}{n_0} \tag{6}$$

where MS_c is the between-cluster mean squares, MS_e, the within-cluster mean squares and n_0 a weighted mean cluster size. The full procedure including some interval estimation methods for ρ are described elsewhere [5]. The coefficient of variation results thus from the ratio of the between-cluster variance over the general mean estimate:

$$^3CV_c = \frac{\hat{\sigma}_r}{\hat{\mu}}. \tag{7}$$

3.1.3. Random-effects models for count data.

The random-intercepts model for the intervention effect on the expected number of events μ_{jl} of Y_{jl} in a CRT has the following form:

$$\log(\mu_{jl}) = \log(t_{jl}) + \beta_0 + \beta x_j + \nu_j , \quad (8)$$

where μ_{jl} is the mean of individual l (l = 1,…,n_j) from cluster j (j = 1,…,K); β_0 the log-mean at the control group (x_j = 0), ν_j the random effect of cluster j, $\nu_j \sim N(0, \sigma_c^2)$; β the effect of the intervention, as the log-means (intervention-over-control) relative rate (RR), x_j the intervention group of cluster j, and t_{jl} the length of individual exposure. Note that the cluster variance is produced at the log scale, that is $V(\log(\mu_j)) = \sigma_c^2$. From the first-order Taylor expansion of μ_j around μ, we obtain:

$$V[\log(\mu_j)] \cong V\left[\log(\mu) + \frac{\partial \log(\mu)}{\partial \mu}(\mu_j - \mu)\right]$$

$$V[\log(\mu_j)] \cong \left[\frac{\partial \log(\mu)}{\partial \mu}\right]^2 V(\mu_j)$$

$$\sigma_c^2 \cong \frac{1}{\mu^2} V(\mu_j)$$

$$\sigma_c \cong \frac{\sqrt{V(\mu_j)}}{\mu} \quad (9)$$

i.e. the CV_c is approximately equal to the square root of the variance component of the cluster effect:

$$^4CV_c \cong \sigma_c . \quad (10)$$

Two distributional assumptions will be considered for count data:

1) Poisson distributed counts, $Y_{jl} \sim \text{Poi}(\mu_{jl})$, with variance function $V(Y_{jl}) = \phi v(\mu_{jl}) = \mu_{jl}$ where ϕ is assumed to be 1; i.e. the mean equals the variance, a property also known as equidispersion that rarely holds in real practice.

2) Negative Binomial (NB) distributed counts, Y_{jl} ~ NB(s, μ_{jl}) with a variance function $V(Y_{jl}) = \phi v(\mu_{jl}) = \phi\left(\mu_{jl} + s\mu_{jl}^2\right)$, where ϕ is assumed to be 1 and s is the NB overdispersion parameter [24].

We consider two alternative approaches for parameter estimation of random-effect models:

i) *Maximum-likelihood based methods (Restricted Pseudo Likelihood method in SAS GLIMMIX)*. Estimates of the model parameters can be obtained by solving the integrals of the likelihood function over the random-effects. We apply the Taylor's series (linearizations) approximations [25] as implemented in the GLIMMIX procedure in SAS v9 [26] and denote it henceforth as GLMM.

ii) *Bayesian estimation via a Markov chain Monte Carlo algorithm*. For a CRT with count outcome Y_{jl} ~ Poi(μ_{jl}) or Y_{jl} ~ NB(s, μ_{jl}) and a model $\log(\mu_{jl}) = \log(t_{jl}) + \beta_0 + \beta x_j + \nu_j$; $\nu_j \sim N(0, \sigma_c^2)$ as specified in (8), the posterior probabilities $P(\beta_0, \beta, \sigma_c^2, s \mid Y)$ are calculated by updating the likelihood $f(Y \mid \beta_0, \beta, \sigma_c^2, s)$ with the prior $P(\beta_0, \beta, \sigma_c^2, s)$ using Markov chain Monte Carlo simulation (MCMC) in the WinBugs Software v1.4 [8].

3.2. Interval estimation of the between-cluster coefficient of variation CV_c

3.2.1. Bootstrap Confidence Intervals.

Bootstrapping is a set of resampling simulation techniques that provide accuracy measures to statistics when their parametrical assumptions seem questionable. For a detailed discussion of the topic, particularly applied to medical statistics we refer to [27]. The method has been also applied in the context of clustered data [6]. In this paper we apply the non parametric bootstrap, with CI obtained from the bootstrap distribution of a large number of re-samples, according to the following algorithm:

1. Sample K clusters randomly with replacement from the original dataset.
2. Calculate the CV_c with one of the methods above.

3. Repeat 1 and 2 a large number of times, to obtain an estimate of the bootstrap distribution.

The CI can be calculated by a number of methods [27]. We apply the non-pivotal percentile method, a technique that uses the $\alpha/2$ and the $100(1 - \alpha/2)$ percentiles of the bootstrap distribution as the lower and upper confidence limits respectively. Its continued popularity among practitioners compared to other non parametric bootstraps is owed to its simplicity and that it is transformation respecting; i.e. when applied to transformed statistics, the back transformed limits to the original scale provide identical limits to those yielded by the untransformed statistics.

3.2.2. Bayesian posterior credible intervals.

The MCMC provides the marginal posterior distribution of σ_c from which the interval limits are obtained as the $\alpha/2$ and $100(1 - \alpha/2)$ percentiles. A comprehensive discussion regarding the choice of (informative/uninformative) priors for between-cluster variation parameters can be found elsewhere [8, 9]

3.2.3. Confidence intervals of generalized pivots.

The concept of generalized pivots, generalized CI and generalized P-values has been developed for a variety of statistics of practical importance where the standard solutions for CI and hypothesis testing may not exist [16, 17, 28, 29]. The method consists of generating a pivotal function of a statistics of interest, with a distribution free of unknown parameters.

Consider model (5) let us define a pivotal quantity for σ_e^2 based on the ANOVA elements and properties as outlined in [16]:

$$V = SS_e \big/ \sigma_e^2 \sim \chi^2_{N-K} \tag{11}$$

with SS_e as the within-cluster sum of squares. The pivot $R_{\sigma_e^2}$ is hence defined as:

$$R_{\sigma_e^2} = \frac{SS_e}{\chi^2_{N-K}}. \tag{12}$$

Assuming the balanced design, the general mean is $\hat{\mu} = \overline{X}_{..} = \sum_{l=1}^{n} \overline{X}_{j.} / K$ as defined above. The mean of cluster *j* are as follows:

$$\overline{X}_{j.} = \sum_{l=1}^{n} X_{jl} / n \sim N\left(\mu, \sigma_e^2 / n + \sigma_r^2\right).$$

Since

$$Q = \frac{\sum_{l=1}^{n} \left(\overline{X}_{j.} - \overline{X}_{..}\right)^2}{\sigma_e^2 / n + \sigma_r^2} = \frac{SS_r}{\sigma_e^2 / n + \sigma_r^2} \sim \chi^2_{K-1}, \tag{13}$$

and solving (13) for σ_r^2 and replacing σ_e^2 by $R_{\sigma_e^2}$, the pivot of σ_r^2 is

$$R_{\sigma_r^2} = \left(\frac{SS_r}{\chi^2_{K-1}} - \frac{SS_e}{\chi^2_{N-K}} \right) \frac{1}{n}. \tag{14}$$

with SS_r as the between-cluster sum of squares. The pivotal quantity of μ is:

$$R_{\mu} = \overline{X}_{..} - \frac{Z}{\sqrt{K / \left(R_{\sigma_e^2} / n + R_{\sigma_r^2}\right)}} \tag{15}$$

where $Z = (\overline{X}_{..} - \mu)\sqrt{K / \left(\sigma_e^2 / n + \sigma_r^2\right)} \sim \mathrm{N}(0, 1)$.

Finally, the pivot for the between-cluster coefficient of variation is:

$$R_{CV} = \frac{\sqrt{R_{\sigma_r^2}}}{R_{\mu}} \tag{16}$$

The computing algorithm is the following:

1. Compute observed versions of SS_e, SS_r and $\overline{X}_{..}$
2. Generate $V \sim \chi^2_{N-K}$, $Q \sim \chi^2_{K-1}$ and $Z \sim$ N(0, 1).
3. Compute $R_{\sigma_e^2}$ from (12), $R_{\sigma_r^2}$ from (14) and R_{μ} from (15).
4. Compute R_{CV} from (16).
5. Repeat 2 – 4 a large number of times, to obtain the sampling distribution of R_{CV}.

The $\alpha/2$ and $100(1 - \alpha/2)$ percentiles of the distribution of R_{CV} would correspond to the lower and upper bounds of R_{CV}. A version of the pivot R_{CV} for unbalanced clusters can be also applied [16].

3.3. Simulation

To assess the methods' performances, random data were generated for three sample sizes: K = (10, 20, 40 total number of clusters), four levels of clustering (σ_c = 0, 0.10, 0.25 and 0.40) and a fixed cluster size of 30 individuals per cluster. A different follow-up time per individual was assumed, being sampled (t_{jl}) from a negative skewed distribution similar to the one observed in the motivating example above: skewness -1.4, mean 290 and s.d. 100, through a power transformation: $t_{jl} = 80(x_{jl}^{1/4})$ where $x \sim$ N(200,100). The control-group event rate θ was set at 5/365 (events per days at risk), and a protective efficacy of 30% was assumed implying a RR of $\exp(\beta) = 0.70$. A cluster effect δ_j was set to act multiplicatively on the mean and whose logarithm was normally distributed with mean 0 and s.d. = σ_c. The number of events Y_{jl} were produced from a NB distribution with mean $\theta t_{jl}\delta_j$ and $\theta t_{jl}\delta_j \exp(\beta)$ for control and intervention clusters respectively and a fixed overdispersion of s = 0.5. Five hundred datasets were generated for each of the 3 × 4 combinations of the defined parameters using different seed numbers.

3.4. Implementation

The CV_c was estimated for each generated dataset by: i) the cluster-level rates method following equations (2) – (4) (CL); ii) the ANOVA method outlined in (6) and (7) and implemented in PROC MIXED of SAS v9.1 [30]; iii) GLMM as implemented in the GLIMMIX procedure of SAS v9.1 [26] following model (8) and assuming a NB distribution.

A Bayesian hierarchical model (Bayes-HM) specified according to model (8) was applied to a random subset of 200 datasets (the high computational demands precluded analysing all 500 datasets by this method). The outcome Y was assumed NB distributed, uninformative priors were used: $\beta_0 \sim N(0, 10^6)$, $\beta \sim N(0, 10^6)$, $\sigma_c^2 \sim IG(0.001, 0.001)$, $s \sim \Gamma(0.001,0.001)$ (IG=Inverse Gamma distribution). A SAS-WinBugs interface was written to analyse the replicate datasets per arrangement in SAS. Model convergence was previously assessed in WinBugs by running two chains with dispersed initial values throughout the parameter space and comparing the between and within chain variation in sample datasets for each of the 12 situations. Convergence was achieved before 5,000 iterations, but 15,000, 10,000 and 7,000 iterations after 1,000 burn-in were implemented in the interface for K = 10, 20 and 40 respectively. The posterior 2.5% and 97.5% quantiles are reported as the intervals (CI for simplicity) and the median as the point estimate.

The 500 datasets were used for the bootstrap method. One thousand bootstraps were run per dataset. Re-sampling was applied at the cluster level, retaining the observations of all subjects in the re-sampled clusters as recommended for cluster designs [6]. The CV_c was then computed for each bootstrap sample by methods i) – iii). The 95% CI's were finally obtained as the 2.5 and 97.5 percentiles of the bootstrap distribution.

The generalized CIs were computed following the computing algorithm outlined in section 3.2.3. 2500 random values for the variates $V \sim \chi^2_{N-K}$, $Q \sim \chi^2_{K-1}$ and $Z \sim N(0, 1)$ were generated for each of the 500 datasets. The 95% confidence limits were calculated as the 2.5 and 97.5 quantiles of the sampling distribution of the pivot R_{CV}.

The point estimation methods were compared in terms of the bias distribution defined as the difference between the underlying CV_c and the observed value. The Interval estimation methods were compared through: 1) coverage probabilities, estimated as the proportion of intervals containing the true CV_c and 2) The interval width (mean and c.v.) as the difference between the upper and lower limits. The programs for data simulations and analysis were written in SAS v9.1 and WinBugs v1.4.

4. RESULTS

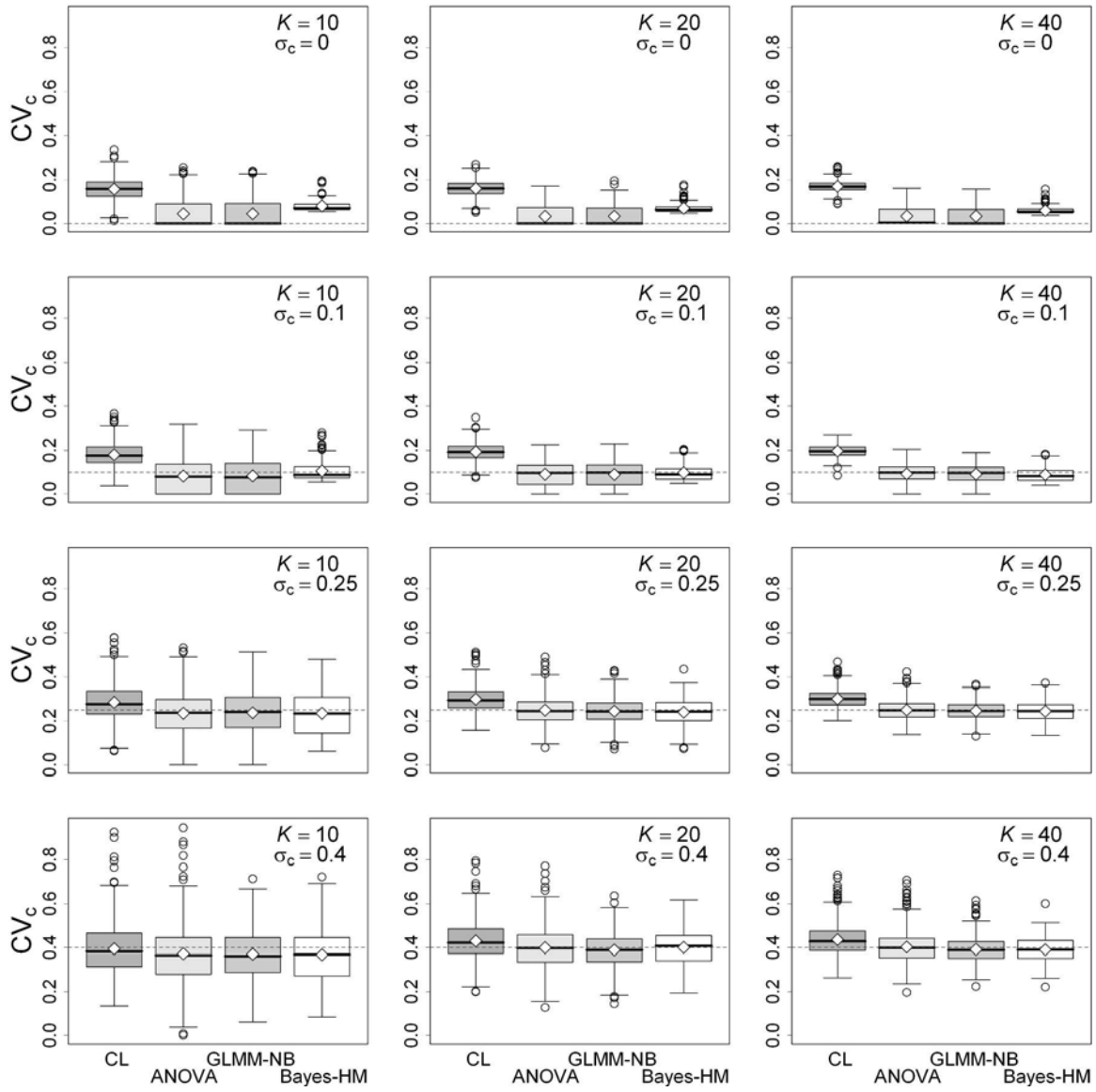
Simulations

Point estimation

The distribution of the point estimates of CV_c by the four corresponding methods is displayed in Figure IV.1. A clear distinction between the cluster-level and the individual-level methods can be appreciated. The CL approach markedly overestimated CV_c particularly when $\sigma_c \leq 0.25$. From the individual-level methods, Bayes-HM showed upward biased estimates when $\sigma_c = 0$ although visibly lower bias than CL. ANOVA and GLMM yielded similar results with CV_c estimates around the expected σ_c. All the methods seem to slightly underestimate CV_c in small sample size and high clustering ($K = 10$, $\sigma_c = 0.40$).

The anticipated effects of K and σ_c on the level of variation of CV_c point estimates were observed. High dispersed estimates were related to either high between-cluster variability or small samples, while more stable estimates were associated to large samples or non correlated data. However, the underlying level of between-cluster variance caused more instability than K, as rather unstable estimates were obtained by all the methods when σ_c was 0.4. Among the four methods, Bayes-HM provided CV_c estimates with a visibly lower variance when $\sigma_c \leq 0.10$ and similar to GLMM when $\sigma_c \geq 0.25$. ANOVA reported similar efficiency than GLMM but somewhat lower when $\sigma_c \geq 0.25$ (Figure IV.1).

Figure IV.1: Between-cluster coefficient of variation (CV_c) by 4 point estimating methods, applied to 200 simulated datasets per combination of number of clusters (K), between-cluster variation (σ_c).

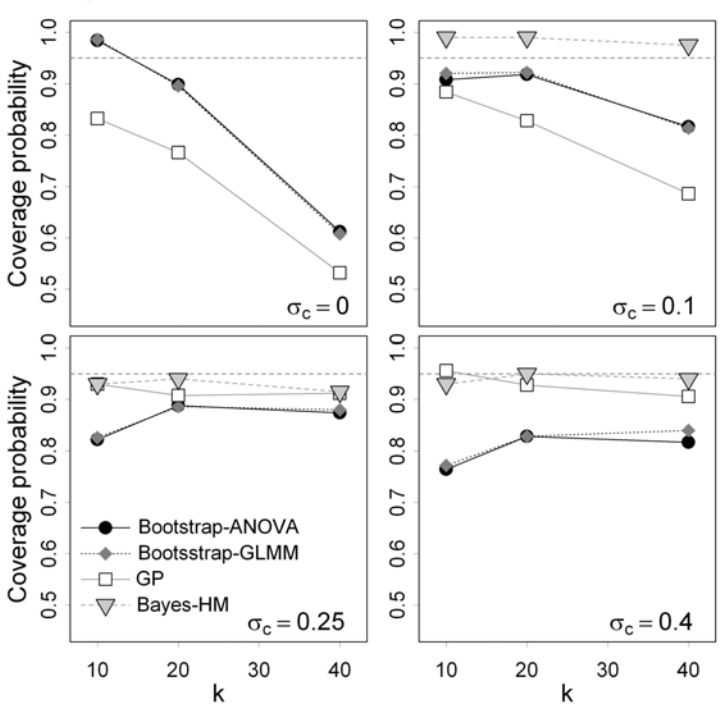


Interval estimation

Figure IV.2 depicts the coverage proportions of the interval estimating methods: the percentile bootstrap applied to the ANOVA (bootstrap-ANOVA) and GLMM (bootstrap-GLMM) point estimates, the generalized pivot CI (GP) and the coverage of the Bayesian credible region. Results from CL are not given because of its highly biased point estimates.

Only Bayes-HM's credible region presented coverage around the nominal 95%, except when $\sigma_c = 0$, the scenario with reported upward bias. Bayes-HM's coverage at $\sigma_c = 0$ was hence 0 regardless of K. Bootstrap-ANOVA and bootstrap-GLMM performed almost identically in all the settings, providing in general inadequate coverage. GP yielded even lower coverage than the bootstraps when $\sigma_c \leq 0.10$, but around nominal when $\sigma_c = 0.40$. Contrary to our expectations, when $\sigma_c \leq 0.10$, the coverage of both bootstraps and that of GP, diminished with the increase of sample size. This result is explained by an upwardly biased sampling (bootstrap or pivot) distribution, whose centre was generally located above the expected σ_c when $\sigma_c \leq 0.10$ (Table IV.1). A reduction of the interval width when K was increased, reduced the chance that the interval included the true σ_c (Table IV.1).

Figure IV.2: Coverage proportions of CV_c interval estimating methods at different: between-cluster variations σ_c and total number of clusters K.



The bootstrap yielded on average the narrowest intervals but also very variable ones. Bayes-HM's interval widths were narrower than GP and approached bootstrap-ANOVA and bootstrap-GLMM in large samples. Bayes-HM had more stable widths than the bootstraps (Table IV.1). With respect to σ_c and K, the interval widths behaved similarly in all the methods. Wider intervals were obtained with high σ_c's or small K, while narrower CI were related to small σ_c's and large K.

Table IV.1: Centers of the sampling distributions across replicate datasets of four corresponding methods for interval estimation of the between-cluster coefficient of variation (CV_c) and mean widths (and cv) of 95% CIs at different sample sizes (K) and level of clustering (σ_c).

σ_c	K	Method	Centre of the Sampling† Distribution		Interval Width	
			Mean	(Q1 , Q3)	Mean	cv
0	**10**	Bootstrap-ANOVA	0.07	(0.03, 0.11)	0.16	40.4
		Bootstrap-GLMM	0.07	(0.03, 0.11)	0.16	40.3
		Generalized Pivot	0.12	(0.07, 0.17)	0.31	28.3
		Bayes-HM	0.08	(0.06, 0.09)	0.22	23.1
	20	Bootstrap-ANOVA	0.08	(0.05, 0.11)	0.14	23.5
		Bootstrap-GLMM	0.08	(0.05, 0.11)	0.15	23.9
		Generalized Pivot	0.11	(0.06, 0.14)	0.22	19.0
		Bayes-HM	0.07	(0.06, 0.08)	0.14	22.2
	40	Bootstrap-ANOVA	0.09	(0.07, 0.11)	0.12	17.7
		Bootstrap-GLMM	0.09	(0.07, 0.11)	0.12	17.5
		Generalized Pivot	0.11	(0.09, 0.14)	0.17	11.7
		Bayes-HM	0.06	(0.05, 0.07)	0.10	20.1
0.1	**10**	Bootstrap-ANOVA	0.11	(0.06, 0.15)	0.20	33.4
		Bootstrap-GLMM	0.11	(0.06, 0.15)	0.20	33.8
		Generalized Pivot	0.16	(0.10, 0.21)	0.34	23.7
		Bayes-HM	0.11	(0.07, 0.13)	0.26	28.7
	20	Bootstrap-ANOVA	0.12	(0.09, 0.15)	0.17	20.9
		Bootstrap-GLMM	0.12	(0.09, 0.15)	0.17	20.9
		Generalized Pivot	0.15	(0.12, 0.19)	0.24	12.7
		Bayes-HM	0.10	(0.07, 0.12)	0.18	21.3
	40	Bootstrap-ANOVA	0.13	(0.11, 0.15)	0.12	20.4
		Bootstrap-GLMM	0.13	(0.11, 0.15)	0.12	20.3
		Generalized Pivot	0.15	(0.13, 0.18)	0.16	8.8
		Bayes-HM	0.09	(0.06, 0.11)	0.13	18.7
0.25	**10**	Bootstrap-ANOVA	0.24	(0.17, 0.30)	0.30	27.4
		Bootstrap-GLMM	0.24	(0.18, 0.31)	0.30	27.0
		Generalized Pivot	0.30	(0.24, 0.37)	0.48	26.3
		Bayes-HM	0.23	(0.14, 0.31)	0.41	23.7
	20	Bootstrap-ANOVA	0.26	(0.22, 0.29)	0.19	28.4
		Bootstrap-GLMM	0.26	(0.22, 0.29)	0.18	21.9
		Generalized Pivot	0.29	(0.25, 0.33)	0.29	15.6
		Bayes-HM	0.24	(0.20, 0.28)	0.26	10.0
	40	Bootstrap-ANOVA	0.26	(0.23, 0.29)	0.13	21.8
		Bootstrap-GLMM	0.26	(0.23, 0.29)	0.12	16.8
		Generalized Pivot	0.28	(0.25, 0.31)	0.19	10.9
		Bayes-HM	0.24	(0.21, 0.28)	0.17	7.5
0.4	**10**	Bootstrap-ANOVA	0.36	(0.28, 0.43)	0.39	33.0
		Bootstrap-GLMM	0.37	(0.29, 0.44)	0.38	27.7
		Generalized Pivot	0.44	(0.34, 0.54)	0.70	44.2
		Bayes-HM	0.37	(0.27, 0.45)	0.54	23.4

20	Bootstrap-ANOVA	0.40	(0.34,	0.45)	0.25	33.3
	Bootstrap-GLMM	0.39	(0.34,	0.44)	0.23	22.4
	Generalized Pivot	0.44	(0.38,	0.50)	0.42	26.3
	Bayes-HM	0.40	(0.34,	0.45)	0.33	14.1
40	Bootstrap-ANOVA	0.40	(0.35,	0.44)	0.23	59.8
	Bootstrap-GLMM	0.40	(0.36,	0.43)	0.16	17.7
	Generalized Pivot	0.43	(0.38,	0.47)	0.27	19.8
	Bayes-HM	0.39	(0.35,	0.43)	0.22	10.4

† *Bootstrap distribution*, *pivot distribution* and *posterior distribution* are referred to for the bootstrap, generalized pivot and Bayes-HM methods respectively.

Analysis of the motivating example

The BoliviaWET trial was powered to estimate a 33% reduction in the diarrhoea incidence rate, assuming 5 episodes per child per year in the control group. Because no prior data existed regarding the extent of between-cluster variation in the study site, sample size calculations were evaluated assuming a range of 0.1 – 0.25 of CV_c from similar community intervention trials [13]. The sample size calculation suggested that at least 18 communities with 10 persons-year of observation per community were sufficient to estimate the desired effect, with a power of 80%, a significance of 0.05 and assuming a CV_c = 0.20. Anticipating a drop-out of at least 2 communities and possible individual drop-outs, the final sample size was adjusted to 22 communities with 30 persons-year of observation [21] .

The estimation of CV_c after data collection, by the point and interval estimating methods is summarized in Table IV.2. For Bayes-HM, 45,000 iterations after 2,000 burn-in were applied.

Two situations were assessed to investigate the effect of overdispersion on CV_c:

i) Specification of Poisson and NB distributions for the GLMM and Bayes-HM analyses.
ii) Data were analyzed with or without the exclusion of outlier observations. Outliers were defined as those having the Pearson's standardized residuals greater than |2.5| for the model with the best fit. This left a remainder of 691 children (Table IV.2).

Complete dataset

The ratio of observed and expected variation in the model ϕ, suggests that the NB model has a substantial better fit than the Poisson models. Estimates of the between-cluster standard deviation and the general mean are reported next. Log scaled values are presented for Poisson and NB analyses, while incidence rate-scaled for the other methods. The resulting CV_c point estimates are later provided.

The cluster-level approach produced a high CV_c point estimate consistent with the simulations results, in which CL visibly overestimated the true value. Note that GLMM and Bayes-HM fits, assuming Poisson errors, produced similarly elevated CV_c, greater than the overdispersion-corrected Poisson and the NB estimates. This suggests that the Poisson assumption may also lead to upwardly-biased CV_c, due to a reallocation of the outcome overdispersion to the between-cluster variance.

GLMM models gave similar results to their Bayesian counterparts. In contrast to what was found during the simulations (Figure IV.1), GLMM with NB errors gave different results from the ANOVA. The ANOVA was the only method that produced $CV_c = 0$, due to truncation of the negative variance component, resulting from a negative difference between MS_c and MS_e (see equation (6)). This was because highly influential observations inflated the MS_e. When the outliers were excluded, a substantial 18.7% of between-cluster variation was obtained by this method.

The CV_c 95% CIs were broad for all methods. The two approaches based on the one-way random-effect model (the bootstrap-ANOVA and GP) provided 0 as the lower limit. The bootstrap-GLMM-Poisson produced narrow intervals, potentially biased and therefore with a higher risk of not including the true CV_c. The Bayes-HM, which showed the best performance during the simulations, yielded also wide intervals.

Outliers excluded

Exclusion of outliers improved model fit for Poisson models and gave almost perfect fit when NB errors were assumed. Both CV_c point estimates and CI widths of all methods were reduced, compared to the analysis of the full dataset, except for the ANOVA CV_c as highlighted before. The point estimates became more similar across the methods and the pattern of interval widths closely resembled the one in the simulation setting to which this trial best matched (Table IV.2).

The evaluation of this example supplied suggestive insight regarding the influence of overdispersion on the between-cluster variance estimation. First, potential upward bias might be expected if overdispersion is not accounted for; i.e., the extra Poisson variation may be artificially allocated to the between-cluster variance. Second, ANOVA-based approaches (including GP) might not be robust to the presence of extreme observations for the random-effects estimation being susceptible to produce anomalous results. Third, extreme observations may similarly influence the NB between-cluster variance although in less extent than the Poisson models. In case of bias, they could be, however, considered conservative estimates.

Table IV.2: Point and Interval estimation of the between-cluster coefficient of variation (CV_c) of the BoliviaWET trial with and without the exclusion of outliers, and assuming Poisson or Negative Binomial distributed counts.

Complete dataset, N=725

Parameter	Cluster-level-Bootstrap	ANOVA-Bootstrap	General. Pivots	Random-effects models for count data				
				Poisson			Negative Binomial	
				GLMM-Bootstrap	GLMM2†-Bootstrap	Bayes-HM	GLMM-Bootstrap	Bayes-HM
ϕ	-	-	-	4.74	4.95	-	1.28	-
$\sigma_r \mid \sigma_c$	0.0033	0.0000	0.0023	0.324	0.195	0.330	0.270	0.271
μ	0.0107	0.0151	0.0150	-	-	-	-	-
CV_c (%)	30.6	0.0	15.4	32.4	19.5	33.0	27.0	27.1
CI of CV_c	(20.6, 42.1)	(0.0, 46.4)	(0.0, 55.2)	(24.0, 39.8)	(11.9, 35.0)	(23.3, 48.7)	(16.9, 40.7)	(10.9, 46.4)
CI width	21.4	46.4	55.2	15.8	23.1	25.4	23.79	35.5

Outliers excluded, N=691

Parameter	Cluster-level-Bootstrap	ANOVA-Bootstrap	General. Pivots	Random-effects models for count data				
				Poisson			Negative Binomial	
				GLMM-Bootstrap	GLMM2†-Bootstrap	Bayes-HM	GLMM-Bootstrap	Bayes-HM
ϕ	-	-	-	2.62	2.66	-	1.02	-
$\sigma_r \mid \sigma_c$	0.0023	0.0018	0.0019	0.262	0.195	0.163	0.195	0.163
μ	0.0088	0.0096	0.0096	-	-	-	-	
CV_c (%)	26.3	18.7	19.9	26.2	19.5	16.3	19.5	16.3
CI of CV_c	(18.5, 34.0)	(12.7, 29.0)	(2.2, 35.7)	(20.4, 32.0)	(11.9, 35.0)	(17.8, 40.0)	(13.7, 29.9)	(3.8, 33.2)
CI width	15.5	16.2	33.5	11.6	23.1	22.3	16.216	29.4

ϕ = Overdispersion parameter, estimated as the generalized Pearson chi-square statistics

σ_r = outcome scaled between-cluster standard deviation. (reported for the non GLMM or Bayesian models).

σ_c =log-scaled between-cluster standard deviation. (reported for GLMM models and Bayes-HM).

μ=Estimate of the general mean.

† GLMM2 = GLMM with standard errors corrected (inflated) by ϕ

5. DISCUSSION

Reporting ρ or CV_c estimates and the computational details are important in CRTs. They facilitate interpretation, and provide information for the design of further trials [31]. In this paper we considered the between-cluster coefficient of variation (CV_c) as the measure to assess clustering in CRTs with overdispersed counts. Point and interval estimation methods of CV_c were studied via simulation under clustering level and sample size conditions similar to those of community-randomized trials.

The CL approach illustrated in [13] for sample size calculation can substantially overestimate the true between-cluster variance in overdispersed counts when the true $CV_c \leq 0.25$. Bayes-HM also showed upward bias in settings without clustering ($\sigma_c = 0$), and similar bias with $\sigma_c = 0.05$ [23], but proved unbiased and efficient when $\sigma_c \geq 0.10$. We may therefore expect the medians of the posterior distribution of σ_c to over-report the clustering level when the between-cluster variation is lower than 10%. However, this bias is less than that of the CL approach and may be considered conservative rather than extreme.

In general, ANOVA and GLMM behaved similarly well regarding CV_c point estimation, although ANOVA generated slightly less efficient estimates in settings with moderate to highly correlated data ($\sigma_c \geq 0.25$). In addition, the efficiency of both methods was seen to decrease in simulations with greater overdispersion (results not shown). In the analysis of the BoliviaWET data, ANOVA, unlike the other methods, suggested there was no between-cluster variation. Additional simulations in which a few observations were replaced by extreme values similar to the ones observed in the BoliviaWET data confirmed that this method can be markedly affected by influential data points.

In terms of interval estimation, the Bayesian credible region had the best performance among the methods studied. Its only disadvantage was related to the observed bias in the posterior distribution when $\sigma_c = 0$. Otherwise, Bayes-HM provided coverage

around the nominal 95% in all settings, and interval widths intermediate between the bootstraps' and the GP's.

Besides Bayes-HM, GP was the only attaining close to nominal coverage but just when σ_c = 0.40, although with rather wide intervals. The poor performance of the percentile bootstrap (applied both to ANOVA and GLMM) and the GP methods, was due to their corresponding bootstrap/pivot distributions being centred away from the expected parameter value when $\sigma_c \leq 0.10$. The interval limits extracted from the percentiles of such distributions are clearly misleading as long as the underlying between-cluster variability is small. Other authors have commented on the percentile bootstrap low coverage and potential bias, proposing alternative procedures or improvements [6, 27]. We assessed this method nonetheless because of its high popularity among practitioners; note that it is implemented in Stata along with other conventional bootstraps.

The GP approach has been successfully applied in the context of other quantities whose sampling distributions may be unknown [16, 17]. We were able to reproduce the very satisfactory findings reported by others when validating our implementation tools in simulated clustered data with normal distribution, but were unable to replicate those findings in clustered negative binomial data. In this sense, a number of normalising transformations were considered and evaluated [32-34]. None of them gave a satisfactory approximation to normality owing to the nature of the individual rates (NB counts/time); small numbers mostly below 0.1, highly skewed and with a prominent mode at 0. The *arcsin* transformation was the one that best approximated the rates to normality but still showed a consistent asymmetry due to the substantial number of zeros. An additional disadvantage is that most transformations are not transformation respecting, that is, the back conversion of the mean and intervals will not correspond to the ones in the original scale. The back transformation will require in consequence a bias correction which in some cases, depending on the transformation, may not be straightforward.

For point estimation of CV_c with overdispersed count data, we consequently recommend GLMM and Bayes-HM assuming NB distribution, with the former

overcoming the conservative bias of the latter in low clustering settings. Point estimation by those methods is based on the extent to which the approximation $CV_c \cong \sigma_c$ holds. To assess this, we considered ANOVA a comparison method, because of its intuitive way of obtaining CV_c on the outcome scale (σ_r/μ). Note that for σ_r point estimation, no individual-level distributional assumption is necessary [22]. ANOVA gave CV_c values similar to those of σ_c by GLMM, with a correlation greater than 0.92 and a change in $\hat{\sigma}_c$ per unit of change in CV_c close to 1 (regression coefficient 0.96), indicative of the 1 to 1 relationship. The two approaches tend to differ however, as the underlying σ_c becomes high, where ANOVA began to report lower estimating efficiency.

Interval estimation of CV_c is a more complex issue. Estimating methods may be based on a series of assumptions than may be difficult to fulfil in real practice or impossible to prove. We considered, for instance, the random effects to be normally distributed. The influence of the misspecification of such distribution has been extensively studied [35, 36] and the maximum-likelihood variance estimates were found to be heavily biased if the underlying distribution is not normal. As the random-effect variances are the only tool to assess the variability of the underlying random-effect distribution, biased estimates due to misspecified distributions will not allow for assessing the validity of fixed effects structure [36]. Bayes-HM through MCMC, and some hierarchical models provide the chance of specifying distributions different than the normal [8, 37]. The use of prior information may be considered also an advantage, provided reports are available of between-cluster variation in similar studies. Other issue regards the difficulty of testing for normality in settings with small number of clusters (community randomized trials). Note that sample sizes required to estimate the intervention effect, are generally smaller than the ones required for appropriate random-effects variance estimation. Considering such implications, we believe Bayes-HM is a reasonable choice for CV_c interval estimation.

This is probably the first study that assessed via simulation methods for point and interval estimation of CV_c, in situations similar to community randomized trials. We assumed overdispersed counts and studied methods attractive among practitioners some already existent in standard statistical software or easy to implement. We

propose CV_c rather than ρ for clustered count data because of the straightforwardness in its calculation. As illustrated already, CV_c values are $\cong \sigma_{c,}$ while estimating ρ would imply i) the conversion of σ_c^2 to the outcome scale and ii) the estimation of residual variance which may depend on the level of overdispersion. We did not investigate alternative bootstrap techniques (e.g. bias-corrected, bias-corrected-accelerated, bootstrap-t) nor extension or modifications of them. Random-effects estimation from more complex designs (e.g. pair-matched, stratified, repeated cross-sectional) was not considered. There is still a need to assess the methods performance on situations with cluster-size imbalance, to fit other extra-Poisson models, and the effect of adjusting for confounders.

Acknowledgments

The BoliviaWET trial was funded by the National Institutes of Health, award number R01AI50087-01IH. Gonzalo Durán Pacheco is in receipt of a stipend from the Stipendiumkommission of the Amt für Ausbildungsbeiträge of the Canton of Basel, Switzerland.

6. REFERENCES

1. Klar N, Donner A. Current and future challenges in the design and analysis of cluster randomization trials. *Stat.Med.* 2001; **20**(24):3729-3740.

2. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. *Am.J.Public Health* 2004; **94**(3):423-432.

3. Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials and Statistics in Medicine. *Stat.Med* 2007; **26**(1):2-19.

4. Donner A, Wells G. A comparison of confidence interval methods for the intraclass correlation coefficient. *Biometrics* 1986; **42**(2):401-412.

5. Ukoumunne OC. A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. *Stat.Med.* 2002; **21**(24):3757-3774.

6. Ukoumunne OC, Davison AC, Gulliford MC, Chinn S. Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient. *Stat.Med.* 2003; **22**(24):3805-3821.

7. Zou G, Donner A. Confidence interval estimation of the intraclass correlation coefficient for binary outcome data. *Biometrics* 2004; **60**(3):807-811.

8. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Stat.Med.* 2001; **20**(3):453-472.

9. Turner RM, Thompson SG, Spiegelhalter DJ. Prior distributions for the intracluster correlation coefficient, based on multiple previous estimates, and their application in cluster randomized trials. *Clin.Trials* 2005; **2**(2):108-118.

10. Turner RM, Omar RZ, Thompson SG. Constructing intervals for the intracluster correlation coefficient using Bayesian modelling, and application in cluster randomized trials. *Stat.Med.* 2006; **25**(9):1443-1456.

11. Ridout MS, Demetrio CGB, Firth D. Estimating intraclass correlation for binary data. *Biometrics* 1999; **55**(1):137-148.

12. Lui KJ, Cumberland WG, Kuo L. An interval estimate for the intraclass correlation in beta-binomial sampling. *Biometrics* 1996; **52**(2):412-425.

13. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *Int.J.Epidemiol.* 1999; **28**(2):319-326.

14. Vangel MG. Confidence intervals for a normal coefficient of variation. *American Statistician* 1996; **50**(1):21-26.

15. Wong ACM, Wu J. Small sample asymptotic inference for the coefficient of variation: normal and nonnormal models. *Journal of Statistical Planning and Inference* 2002; **104**(1):73-82.

16. Tian L. Inferences on the within-subject coefficient of variation. *Stat.Med.* 2006; **25**(12):2008-2017.

17. Tian L. On confidence intervals of a common intraclass correlation coefficient. *Stat.Med.* 2005; **24**(21):3311-3318.

18. Sommer B, Marino A, Solarte Y, Salas ML, Dierolf C, Valiente C, Mora D, Rechsteiner R, Setter P, Wirojanagud W, Ajarmeh H, AlHassan A, Wegelin M. SODIS - An emerging water treatment process. *Journal of Water Supply Research and Technology-Aqua* 1997; **46**(3):127-137.

19. Dejung S, Fuentes I, Almanza G, Jarro R, Navarro L, Arias G, Urquieta E, Torrico A, Fenandez W, Iriarte M, Birrer C, Stahel WA, Wegelin M. Effect of solar water disinfection (SODIS) on model microorganisms under improved and field SODIS conditions. *Journal of Water Supply Research and Technology-Aqua* 2007; **56**(4):245-256.

20. Conroy RM, Elmore-Meegan M, Joyce T, McGuigan KG, Barnes J. Solar disinfection of drinking water and diarrhoea in Maasai children: a controlled field trial. *Lancet* 1996; **348**(9043):1695-1697.

21. Mäusezahl, D., Christen, A., Duran-Pacheco, G., Alvarez-Tellez, F., Iriarte, M., Zapata M.E., Cevallos, M., , Hattendorf J., M., Arnold, B., Smith-A T., and Colford, J. M. A cluster-randomized, controlled trial of solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia. 2008.

22. Searle S.R. *Linear Models*. Wiley & Sons: New York, 1997.

23. Duran-Pacheco, G., Hattendorf, J., Colford, J. M., Mäusezahl, D., and Smith, T. Performance of analytical methods for overdispersed counts in cluster randomized trials: sample size, degree of clustering and imbalance. *Statist. Med.* 2009; **28**:2989–3011.

24. Hilbe J.H. *Negative Binomial Regression*. Cambridge University Press, New York: 2007.

25. Rodriguez G, Goldman N. An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. *Journal of the Royal Statistical Society Series A-Statistics in Society* 1995; **158**:73-89.

26. SAS Institute Inc. *The GLIMMIX Procedure*. SAS Institute Inc.: Cary, North Carolina, USA, 2006.

27. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* 2000; **19**(9):1141-1164.

28. Tian L. Interval estimation and hypothesis testing of intraclass correlation coefficients: the generalized variable approach. *Stat.Med.* 2005; **24**(11):1745-1753.

29. Iyer HK, Wang CMJ, Mathew T. Models and confidence intervals for true values in intertaboratory trials. *Journal of the American Statistical Association* 2004; **99**(468):1060-1071.

30. SAS Institute Inc. *SAS/STAT 9.1 user's guide*. SAS institute Inc.: Cary: NC, 2004.

31. Campbell MK, Elbourne DR, Altman DG. The CONSORT statement for cluster randomised trials. *Medicina Clinica* 2005; **125**:28-31.

32. Berry DA. Logarithmic Transformations in Anova. *Biometrics* 1987; **43**(2):439-456.

33. Peltier MR, Wilcox CJ, Sharp DC. Technical note: Application of the Box-Cox data transformation to animal science experiments. *Journal of Animal Science* 1998; **76**(3):847-849.

34. Rate transformations and smoothing (technical report). http://geodacenter.asu.edu/pdf/smoothing_06.pdf. Accessed 10/04/2009.

35. Heagerty PJ, Kurland BF. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 2001; **88**(4):973-985.

36. Litiere S, Alonso A, Molenberghs G. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine* 2008; **27**(16):3125-3144.

37. Lee KJ, Thompson SG. Flexible parametric models for random-effects distributions. *Statistics in Medicine* 2008; **27**(3):418-434.

Chapter V:

Reporting diarrhoea through a vernacular term in Quechua speaking settings of rural Bolivia

Gonzalo Durán Pacheco[1], Andri Christen[1], Ben Arnold[2], Jan Hattendorf[1], Ada Armaza[3], John M. Colford, Jr.[2], Thomas Smith[1], Daniel Mäusezahl[1]

[1] *Department of Public Health and Epidemiology, Swiss Tropical Institute, Basel, Switzerland.*
[2] *School of Public Health, University of California, Berkeley (UCB), California*
[3] *Instituto de Investigaciones Biomédicas, Facultad de Medicina, Universidad de San Simón, Cochabamba Bolivia*

This manuscript has been submitted to: *Journal of Health, Population and Nutrition*

ABSTRACT

This paper describes the association of the vernacular Quechua term *k'echalera* with the symptoms-based standard definition of diarrhoea in rural Bolivian settings. Signs and symptoms of diarrhoea as well as *k'echalera* reports were collected during a cluster randomized trial in rural Bolivia. Reports of *k'echalera* were found to be associated with important changes in stool frequency, consistency and occurrence of blood and mucus. *K'echalera* reports were highly related to three types of watery-stool consistencies from the four applied in field tools. The intermediate *milky rice* stool consistency which fits into the definition of watery stool was not strongly related to *k'echalera*. Mucus in the stool was also associated with *k'echalera* and its occurrence in *k'echalera*-free days accounted for at least 50% of the possible false negatives. Sensitivity and specificity of the term *k'echalera* was estimated by Bayesian methods allowing for both the diarrhoea symptoms and *k'echalera* reports to be subject to diagnosis error. We obtained an average specificity of at least 97% and sensitivity of at least 50%.

Keywords: Diagnosis of diarrhoea; caregiver's reports; Quechua vernacular terms; *K'echalera*; rural Bolivia

1. INTRODUCTION

Based on a common set of signs and symptoms, diarrhoea is defined as the obvious change in the normal stool pattern, characterized by 3 or more watery loose stools in a 24 h period or 1 or more stools with evident presence of blood or mucus (Baqui *et al.* 1991;Jeejeebhoy 1977;Morris *et al.* 1994;Thapar & Sanderson 2004).

Reports of mothers or caregivers are also used and widely accepted for reporting of diarrhoea occurrence in children (Killewo & Smet 1989;Pathela *et al.* 2006;Ruel *et al.* 1997). Vernacular terms are then necessarily employed and morbidity estimates calculated from these. The validity of such reports is based on the observation that people who regularly care for young children are aware of the actual change in the child's normal habits of stool frequency, volume and consistency (Baqui *et al.* 1991; Morris *et al.* 1994). The correspondence between mother-defined and symptom-based definitions may vary across populations and cultures (Baqui, *et al*. 1991).

K'echalera is a generic term widely used in Quechua-speaking settings of South America (from northern Ecuador, to southern Bolivia). It refers to a change in the ordinary stool patterns as a result of an increased volume and frequency of stool with simultaneous change of stool consistency. The term has also been adopted as part of the folk and Criollo language in urban Spanish-speaking areas in Bolivia (Prudencio C.A. 1978) and is used by health and medical staff to assess diarrhoea in rural areas. Eleven specific terms (e.g. *K'echa Pukay*, *K'echa K'ellu*, *K'echa Yuraj*) have been found to classify gastrointestinal illness by colour, odour and frequency of stool, standing *k'echalera* in general for watery and frequent stool (Hobbins 2004).

This report aims at describing the association of the term *k'echalera* with the symptoms-based standard definition and to estimate the sensitivity and specificity of the vernacular definition relative to the international standard.

2. METHODS

Data

We use data from a baseline survey and the fist six-months of the post-intervention follow-up of a recent community randomized trial on solar water disinfection in rural Bolivia (BoliviaWET) (Mäusezahl *et al.* 2009). Weekly and daily diarrhoeal symptoms and occurrence of *k'echalera* were collected for the eight-weeks baseline and the post-intervention follow-up respectively. Mothers or primary caregivers of study participants provided data regarding: number of stools during the last 24 hrs, stool consistency, presence of blood or mucus as well as *k'echalera* occurrence. We identified local foods to use as stool consistency analogs to standardize our measurement in focus group sessions in our study population. We used the Quechua versions of the following analogs to measure stool consistency: *liquid* (water, api), *semi-liquid* (arrope), *intermediate* (milk rice), *semi-solid* (mashed potatoes), *solid* (sausage) (Table V.1).

Data analysis

Descriptive and exploratory

The distribution of diarrhoeal symptoms is compared for days with and without reported *k'echalera*. The correspondence among answers to the questionnaire concerning: number of stools, consistency of stool, presence of blood and mucus, was analysed by a multiple correspondence analysis (MCA) on the Burt matrix (Lebart *et al.* 2000). The association between categories of different variables was simultaneously visualized by a scatter plot of the first two factorial axes. Closeness between categories of different symptoms should be interpreted as association.

Estimating the Sensitivity and Specificity

A variable describing the standard symptom-based definition (*std-diarrhoea*) was defined as the daily passage of at least 3 watery loose stools or at least one stool containing blood or mucus. *K'echalera* reports were contrasted with those of *std-diarrhoea* (Table V.2). We assumed that both *k'echalera* and *std-diarrhoea* are susceptible to diagnostic error. We hypothesize that the report of symptoms may be subject to measurement error depending on how knowledgeable the caregiver is in the child's regular patterns of defecation. In addition, cultural norms when reporting to the field staff may contribute to reporting bias. Since standard methods of calculating diagnostic statistics assume that the "gold standard" method is the truth (an assumption that may not reasonably hold in this analysis), we estimate sensitivity (*Se*) and specificity (*Sp*) using Bayesian methods (Black & Craig 2002;Gustafson 2005), which allow both metrics – *k'echalera* and *std-diarrhoea* – to be measured with error.

Informative (beta distributed) priors for the sensitivity and specificity of *std-diarrhoea* (^{d}Se and ^{d}Sp) were employed. We assumed *std-diarrhoea* to be highly sensitive and specific (mode of ^{d}Se and ^{d}Sp = 0.95) but a 95% chance of being at least 0.8. Provided the high observed specificity (Table V.2) and negative predictive value of *k'echalera*, informative (Beta) priors were used for the sensitivity and specificity of *k'echalera* (^{k}Se and ^{k}Sp). We assumed ^{k}Sp to have a mode = 0.95 but 95% chances of being at least 0.80. More uncertainty was assumed about the knowledge of ^{k}Se, and three priors were assessed:

i) Full uncertainty (uninformative prior: $^{k}Se \sim$ Beta(1,1)).
ii) Vague optimistic prior (mode = 0.7 and 95% chances of being at least 0.3)
iii) Vague pessimistic prior (mode = 0.3 and 95% chances of being at most 0.70).

Finally a prior assuming complete ignorance of the prevalence of diarrhoea (λ) was also evaluated ($\lambda \sim$ Beta (1,1)). Figure V.2 displays the assumed prior uncertainty on ^{d}Se, ^{d}Sp, ^{k}Se and ^{k}Sp.

3. RESULTS

The distribution of the diarrhoeal symptoms is reported in Table V.1 for days with and without *k'echalera* from the pre-intervention study and days with *k'echalera* from the post-intervention follow-up. A day without *k'echalera* was characterized by a median of 1 stool, mostly solid or semisolid (69.8%). Although in much lower proportion, blood and mucus were also reported in days without *k'echalera*. Days with *k'echalera* in the pre-intervention study were characterized by a median of 3 stools during the last 24 hrs, a predominant proportion of watery stool (81.1%), and higher frequency of blood or mucus presence compared to days without *k'echalera*. Watery stool was defined as one that would take the shape of the container (Clasen *et al.* 2007;Ejemot *et al.* 2008).

Table V.1: Distribution of the diarrhoeal symptoms for days with and without *k'echalera* in Baseline and a post-intervention study.

Symptom	Pre-intervention Days without *K'echalera* N = 4071	Pre-intervention *Days with K'echalera* N = 281	Post-intervention *Days with K'echalera* N = 4412
Nr of stools, last 24 hrs: median (Q1; Q3)	1 (1; 2)	3 (2; 3)	3 (2, 4)
Stool consistency: *n (%)*			
Liquid (water)	142 (3.5)	102 (36.3)	2021 (45.8)
Liquid (api[†])	76 (1.9)	48 (17.8)	931 (21.1)
Semi-liquid (arrope[‡])	186 (4.6)	62 (22.1)	912 (20.7)
Intermediate (milk rice)	177 (4.4)	14 (4.9)	249 (5.6)
Watery stool: Total	**581 (14.3)**	**228 (81.1)**	**4113 (93.2)**
Semi-solid (mashed potatoes)	865 (21.3)	24 (8.5)	102 (2.3)
Solid (sausage)	1975 (48.5)	16 (5.7)	6 (0.14)
Solid or semi-solid: Total	**2840 (69.8)**	**40 (14.2)**	**108 (2.5)**
Other	1 (0.02)	1 (0.4)	78 (1.8)
Don't know	649 (15.9)	12 (4.3)	113 (2.6)
Blood in the stool: *n (%)*	51 (1.25)	39 (13.9)	666 (15.1)
Mucus in the stool: *n (%)*	231 (5.7)	97 (34.5)	1965 (44.5)

N, n = nr of days
Pre-intervention data represent once-a-week data
Post-intervention data represent daily data
[†]api: a non-alcoholic thick corn drink
[‡] arrope: a non-alcoholic beverage, quite tick sweet syrup, produced by adding water to *Prosopis flour* (borra).

Among the watery loose stools categories, “milk rice” is equally likely in both days with and without *k’echalera*. Similar patterns were observed in the post-intervention data with a much larger sample size. Here, the proportion of watery stool was higher (93.2%) than in baseline (81.1%), owed to the increase of liquid and decrease of solid and semi-solid consistencies (Table V.1). A characterization of days without *k’echalera* was not provided for the post-intervention period, because data on diarrhoeal symptoms were collected only if *k’echalera* was reported.

Figure V.1: Distribution of the modalities of the diarrhoeal symptoms of the questionnaire and the reports of *k’echalera* in a plane conformed by the 2 first factorial axis of a multiple correspondence analysis.

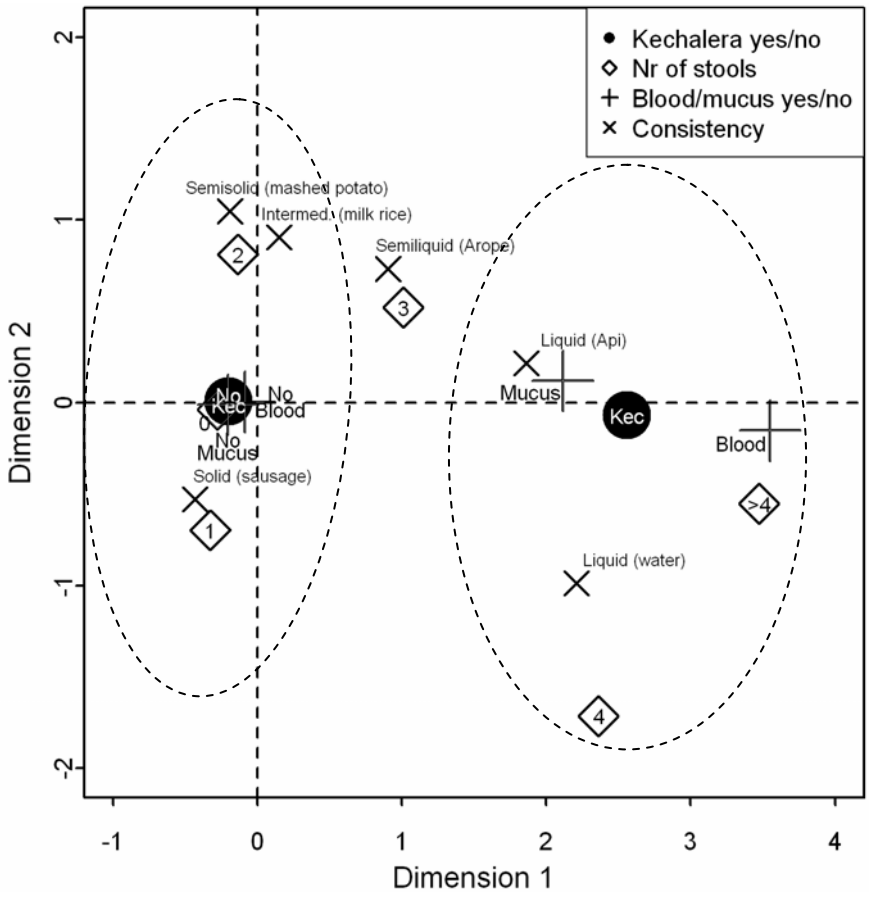


Figure V.1, displays the distribution of the categories of the four diarrhoeal symptoms and the *k’echalera* status in a factorial space obtained by MCA. The figure reflects joint symptoms reported for children on the same day of observation. *K’echalera*, contrasts with *no k’echalera* by being at the centre of the categories that do characterize diarrhoea, i.e.: blood, mucus, the two forms of liquid consistency assessed and high number of stools. This suggests that whenever *k’echalera* was

reported, the diarrhoeal symptoms were reported too. Conversely, *no k'echalera* was reported in the absence of blood, mucus, solid or semisolid stools. Interestingly, three stools per day and semi-liquid stool consistency modalities fall approximately equidistant between the *k'echalera* and *no k'echalera* classifications; this suggests that these symptom categories are where the two classifications begin to overlap. Indeed, from all the semi-liquid reports in days with *k'echalera* (*n*=61), 85.5% were given when ≥ 2 stools were reported (35.5% correspond to 2 stools). Conversely, 95.2% (*n*=183) of the semi-liquid stools in *k'echalera*-free days were reported when ≤ 3 stools were reported (14.0%, 34.9% 40.3% for 3, 2 and 1 stools respectively). The intermediate *milk rice* and semisolid stool consistencies fall closer to days without *k'echalera* because both of them were frequently reported together with 2 stools.

Observed sensitivity and specificity

Table V.2 shows the distribution of the days with *k'echalera* across the combination of diarrhoeal symptoms that make the standard definition *std-diarrhoea*.

Table V.2: Sensitivity and specificity of *k'echalera* reports compared to the standard symptom-based definition of diarrhoea.

	Std-Diarrhoea	
K'echalera	Days with	Days without
Days with	177	100
Days without	315	3434

Assuming that *std-diarrhoea* is the gold standard, the observed sensitivity of *k'echalera* was 36% (177/492). The main reason for a low sensitivity was the large number of false negatives. From the 315 days without *k'echalera* but positive according to *std-diarrhoea*, 104 reported at least 3 watery loose stools, 16 reported at least 1 stool with blood, 168 reported mucus, and 26 both mucus and blood (Table V.3). The reasons for the 100 apparent false positives are also presented in Table V.3. The prevalence calculated following the *std-diarrhoea* definition yields 12.2% (492/4026) while a prevalence following the *k'echalera* definition suggests 6.9% (277/4026).

The observed specificity 97.2% (3434/3534) and negative predictive value 91.2% (3434/3749) were high.

Table V.3: Reasons of false negative and false positive reports of *k'echalera* using the standard symptom-based definition of diarrhoea as gold-standard

	Reported symptom	n (%)
False Negatives	≥3 Watery loose stools, no blood no mucus	104 (33.1)
	≥1 stool with only blood	16 (5.1)
	≥1 stool with only mucus	168 (53.5)
	≥1 stool with both blood and mucus	26 (8.3)
	Missing	1
	Total	315
False Positives	<3 stools, no blood, no mucus	74 (74.0)
	3 solid or semisolid stools (no blood, no mucus)	10 (10.0)
	missing	16 (16.0)
	Total	100

Modelling the sensitivity and specificity

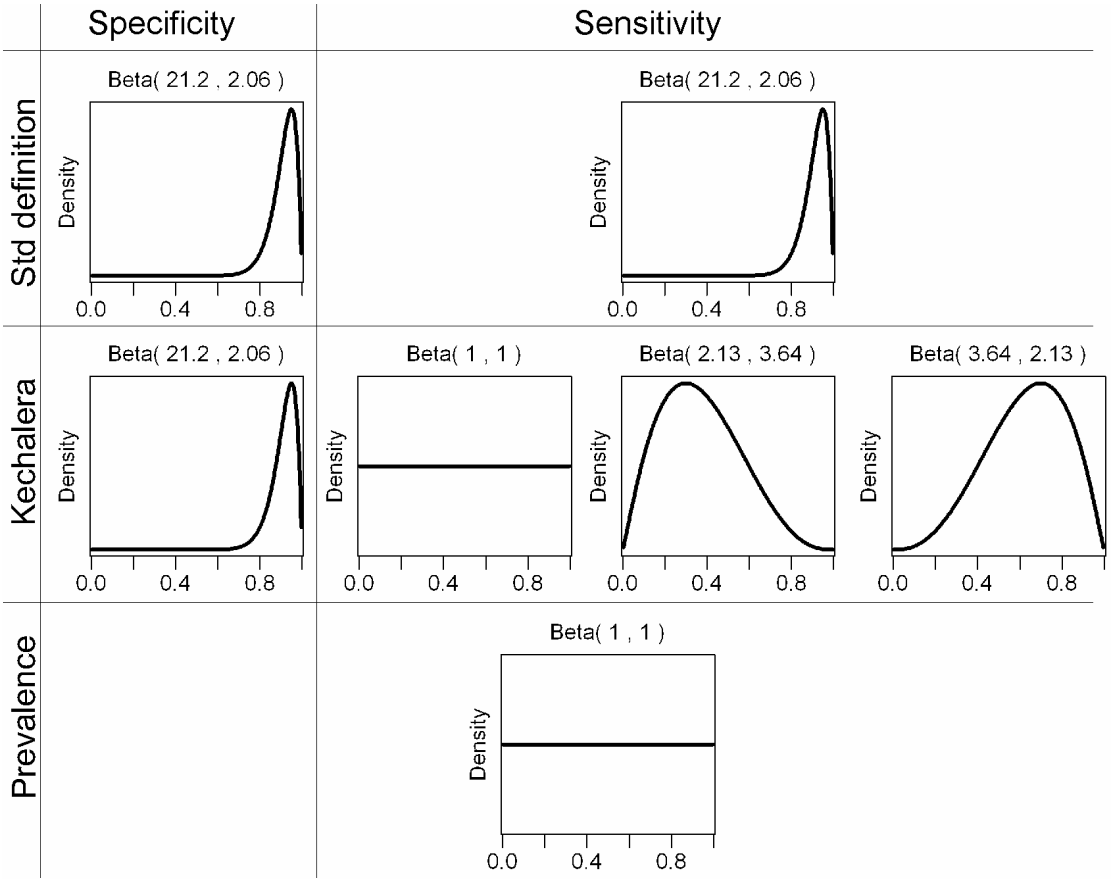
Assuming that both *k'echalera* and *std-diarrhoea* are subject to diagnostic error or recall bias, the sensitivity and specificity estimates using the uncertainty levels displayed in Figure V.2, are presented in Table V.4. Note that we presume to be more certain on the high specificity of *k'echalera* and on the high *Se* and *Sp* of the standard definition.

Table V.4: Estimates of the sensitivity and specificity of *K'echalera* and the standard definition allowing for uncertainty in their reporting accuracy (pre-intervention data)

	Prior for *k'echalera*	**Sensitivity**	**Specificity**
K'echalera	Uninformative	60.8 (38.1; 97.4) †	97.5 (96.8; 98.6)
	Optimistic	61.9 (39,3; 91.7)	97.6 (96.8; 98.6)
	Pessimistic	49.6 (36.1; 77.6)	97.6 (96.8; 98.6)
std-diarrhoea	Uninformative	92.4 (78.2; 98.8)	94.4 (91.4; 98.9)
	Optimistic	92.2 (78.3; 98.8)	94.3 (91.7; 98.6)
	Pessimistic	92.5 (78.4; 98.8)	96.1 (92.7; 99.3)
Prevalence of diarroea	Uninformative	7.7 (4.5; 12.8)	
	Optimistic	7.6 (4.8; 12.4)	
	Pessimistic	9.5 (5.8; 13.3)	

†Posterior median and credible interval

Figure V.2: Prior distributions of the sensitivity and specificity of *k'echalera* and for the functional definition of diarrhoea based on reported symptoms.



Regardless of prior beliefs about the sensitivity of *k'echalera* (uninformative, vaguely optimistic and vaguely pessimistic), ${}^{k}Se$ was always estimated higher than the observed values calculated from Table V.2. Introducing a reasonable level of uncertainty in the report of the *std-diarrhoea* symptoms led to an important increase in ${}^{k}Se$ to 50% with the pessimistic prior and 62% with the optimistic one (Table V.4). ${}^{k}Sp$ was always high. The prevalence of diarrhoea was estimated around 7.7% assuming uninformative and optimistic priors and 9.5% assuming a pessimistic prior for ${}^{k}Se$ (Table V.4).

4. DISCUSSION

We evaluated the meaning of the vernacular term *k'echalera* as a mother/care giver diagnosis of diarrhoea in rural Bolivian settings and compared its reporting to an internationally standardized, symptom-based diarrhoea definition. We found that

caregivers use the term *k'echalera* to reflect a noticeable change in the child's regular defecation patterns characterised by an increase of bowel movement frequency and a change in the stool consistency. A median of 3 watery stools during the last 24 hrs, 81.1% of the stools in days with *k'echalera* had a watery consistency, and a greater proportion of blood and mucus compared with days without *k'echalera.* The proportion of watery stool was confirmed to be greater (93.2%) in *k'echalera* days when measured in the post-intervention data. We found some divergence in the vernacular use of *k'echalera* and the international standard definition of diarrhoea. A *k'echalera* report was strongly associated with liquid and semi-liquid stools that differ clearly from solid stool. However, the *intermediate* stool consistency level (milk-rice-like stool), which fits into the definition of watery loose stool (Clasen *et al.* 2007;Ejemot *et al.* 2008), did not help to discriminate between *k'echalera* and non-*k'echalera*. Blood and mucus in the stool were also positively associated with *k'echalera*. Mucus was reported during days without *k'echalera* in a much lower proportion, but enough to increase appreciably the number of false positives.

These observed reporting differences led to a low sensitivity of the vernacular term compared to the standard symptom-based diarrhoea definition. The reporting differences led principally to false negatives, characterized by episodes with high stool frequency and intermediate consistencies, or days with at least 1 stool with mucus. The specificity and negative predictive value of *k'echalera* were consistently high. A bayesian analysis that allowed for measurement error in both *k'echalera* and the symptom-based definition of diarrhoea (a scenario that we argue more accurately reflects real measurement conditions) increased the vernacular term's sensitivity from 36% to between 50% and 62%.

In addition, we hypothesize that discrepancies between *k'echalera* and the symptoms reports might both be due to two main sources of measurement error: i) perception/detection by the caregiver, influenced by how much time the caregiver spends with the child and how much attention she pays to stool symptoms, and ii) the caregiver reporting to the field staff, influenced by cultural norms, practices and social desirability and the relationship between the caregiver and the field staff. Moreover, we wished to allow *std-diarrhoea* as possibly deviating from the actual changes in defecation patterns in the study setting. In this sense the estimation of the sensitivity

of the term *k'echalera* was done using Bayesian techniques allowing for a reasonable level of uncertainty to the report of symptoms. A higher sensitivity was then obtained and validated through a sensitivity analysis of the priors employed.

Assuming the symptoms-based definition is the gold standard, maternal reports of diarrhoea in different settings yielded higher *Se* estimates than ours in Table V.2. Baqui and colleagues (Baqui *et al.* 1991) actually assumed that the mother's definition is the gold standard. They provide data, however, suggesting that *Se* of the mother's definition compared to the standard is 68% (in line with our 61% estimate using uninformative and vague optimistic priors for ${}^{k}Se$). A study in South Africa (Ferrinho *et al.* 1995) reported even a higher sensitivity of 89% for the mothers' report. However the latter estimate was obtained comparing diarrhoea occurrence over a 1-2 months recall period with the occurrence of symptoms in the same period. In contrast, our study, like others (Baqui *et al.* 1991), compared reports of symptoms and *k'echalera* occurrence corresponding to one day of observation. Thomas *et al* (1989) provided *Se* and *Sp* estimates for mothers' reports of diarrhoea being 79% and 94% respectively. A study in the Philippine island of Cebu (Kalter *et al.* 1991), provided *Se* and *Sp* estimates of maternal symptom-based diagnosis as compared with physicians' diagnosis. The diagnosis of diarrhoea had a sensitivity of 95-97% and a specificity of 80% when based on maternal reports of frequent loose of liquid stools. That suggests that mothers were able to retrospectively report the signs and symptoms of their children accurately for interview-based diagnosis. That *Se* and *Sp* concur with our assumption on the priors for the symptoms-based definition in the Bayesian analysis.

Our crude prevalence estimates fall between 6.9 and 12.2% for *k'echalera* and the symptom-based diarrhoeal reports respectively. This suggests that, in our study setting, mothers do not identify diarrhoea very consistently with the international definition. In contrast to other cultures, in many cases mothers reported the presence of mucus and milk-rice consistency as "normal", what other cultures would report as diarrhoea (Bangladesh (Baqui *et al.* 1991), South Africa (Ferrinho *et al.* 1995), Kenya (Thomas, Neuman Ch G., & Frerichs 1989)). We found a high prevalence of malnourished children, especially wasted children (data not shown). This health status was often accompanied with mal absorption of food and chronic diarrhoea with *milk rice* stool consistency. In addition, the mal absorption of food and the resulting unshaped stool

was often accompanied by mucus a well described physiological phenomenon (Thapar & Sanderson 2004). We presume that such health status was perceived as normal by the mother and reported as day without *k'echalera*.

We believe that the prevalence of diarrhoea lies between the *k'echalera* and *std-diarrhoea* estimates and the reasonable uncertainty assumed during the Bayesian analysis is a good approximation (7.6 – 9.5%). The disadvantage of this approach is that good care should be taken when choosing the priors, since the final estimates may be sensitive to their choice.

Conclusion

In this rural Bolivian population, the term *k'echalera* is used to report a true change in the defecation patterns of children under 5 years. *K'echalera* is strongly associated with the symptoms that are used in the symptom-based standard definition. However, the intermediate (milk-rice) stool consistency and mucus presence, part of the standard definition, were frequently reported in days without *k'echalera* and were responsible for numerous false negative results. We estimated an average sensitivity of *k'echalera* of at least 50% and a specificity of 97% when allowing for uncertainty on both *k'echalera* and the symptoms report. The low sensitivity of *k'echalera* relative to the standard definition may be due, in part, to caregivers perceiving as normal chronic, low-level diarrhoeal symptoms that classify children as diarrhoeic in other settings.

Acknowledgments

We warmly thank the participating families and acknowledge the support of the study communities in Bolivia. We also acknowledge the field staff of the BoliviaWET study for their relentless commitment during data collection. We like to specifically thank Stefan Indengard and Michael Hobbins for their valuable contributions on assessing vernacular terminology of child diarrhoea in rural Bolivia. We are grateful to Ada Armaza for her valuable comments on a previous version of the manuscript. The BoliviaWET trial was funded by the National Institutes of Health, award number R01AI50087-01IH. Gonzalo Durán Pacheco is in receipt of a stipend from the

Stipendiumkommission of the Amt für Ausbildungsbeiträge of the Canton of Basel, Switzerland.

5. REFERENCES

1. Baqui AH, Black RE, Yunus MD *et al.* (1991) Methodological Issues in Diarrheal Diseases Epidemiology - Definition of Diarrheal Episodes.*International Journal of Epidemiology* **20**, p 1057-1063.

2. Jeejeebhoy KN (1977) Symposium on diarrhea. 1. Definition and mechanisms of diarrhea.*CMA Journal* **116**, p 737-739.

3. Morris SS, Cousens SN, Lanata CF, & Kirkwood BR (1994) Diarrhoea--defining the episode.*Int J Epidemiol* **23**, p 617-623.

4. Thapar N & Sanderson IR (2004) Diarrhoea in children: an interface between developing and developed countries.*Lancet* **363**, p 641-653.

5. Killewo JZ & Smet JE (1989) Mother's definition of diarrhoea in a suburban community in Tanzania.*J.Diarrhoeal Dis Res* **7**, p 21-23.

6. Pathela P, Hasan KZ, Roy E *et al.* (2006) Diarrheal illness in a cohort of children 0-2 years of age in rural Bangladesh: I. Incidence and risk factors.*Acta Paediatrica* **95**, p 430-437.

7. Ruel MT, Rivera JA, Santizo MC, Lonnerdal B, & Brown KH (1997) Impact of zinc supplementation on morbidity from diarrhea and respiratory infections among rural Guatemalan children.*Pediatrics* **99**, p 808-813.

8. Prudencio C.A. *Diccionario del cholo ilustrado.* Ojo Publicaciones, La Paz, Bolivia (1978): 95

9. Hobbins, M. Home-based drinking water purification through sunlight: from promotion to health effectiveness. 2004. University of Basel. 214 p. (Dissertation)

10. Mäusezahl D, Christen A, Duran Pacheco G *et al.* (2009) Solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia: A Cluster-Randomized, Controlled Trial.*PLOS Medicine* **6**. doi:10.1371/journal.pmed.1000125.

11. Lebart L, Morineau A, & Piron M (2000) *Statistique exploratoire multidimensionnelle.* 3rd edn. Dunod, Paris.

12. Black MA & Craig BA (2002) Estimating disease prevalence in the absence of a gold standard.*Statistics in Medicine* **21**, p 2653-2669.

13. Gustafson P (2005) The utility of prior information and stratification for parameter estimation with two screening tests but no gold standard.*Statistics in Medicine* **24**, p 1203-1217.

14. Clasen T, Schmidt WP, Rabie T, Roberts I, & Cairncross S (2007) Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis.*BMJ* **334**, p 782. doi:10.1136/bmj.39118.489931.BE

15. Ejemot RJ, Ehiri JE, Meremikwu MM, & Critchley JA (2008) Hand washing for preventing diarrhoea.*Cochrane Database of Systematic Reviews*.

16. Ferrinho P, Ratsaka M, Bellingham A, & Groneveld H (1995) Methodological aspects of a household survey on diarrhoeal diseases in a peri-urban community of South Africa - The problem of defining diarrhoea.*Journal of Tropical Pediatrics* **41**, p 315-317.

17. Thomas JC, Neuman Ch G., & Frerichs R (1989) The effect of misclassification of diarrhoea on estimates of its occurrence, the identification of risk factors and the assessment of prevention efforts.*J.Diarrhoeal Dis Res* **7**, p 63-69.

18. Kalter HD, Gray RH, Black RE, & Gultiano SA (1991) Validation of the Diagnosis of Childhood Morbidity Using Maternal Health Interviews.*International Journal of Epidemiology* **20**, p 193-198.

Chapter VI:

Factors associated with compliance among users of solar water disinfection in rural Bolivia

Andri Christen[1], Gonzalo Durán Pacheco[1], Jan Hattendorf[1], Myriam Cevallos[1], Carlos Morante Navarro[3], Ben Arnold[2], John M. Colford, Jr.[2], Daniel Mäusezahl[1]

[1] *Department of Public Health and Epidemiology, Swiss Tropical Institute, Basel, Switzerland.*
[2] *Division of Epidemiology, School of Public Health, University of California, Berkeley (UCB), California*
[3] *Project Concern International (PCI), Cochabamba, Bolivia*

This manuscript has been submitted to: *BMC Public Health*

Abstract

Background

Diarrhoea is the second leading cause of childhood mortality, with an estimated 1.3 million deaths per year. Promotion of Solar Water Disinfection (SODIS) has been suggested as a strategy for reducing the global burden of diarrhoea by improving the microbiological quality of drinking water. Despite increasing support for the large-scale dissemination of SODIS, there are few reports describing the effectiveness of its implementation. It is, therefore, important to identify and understand the mechanisms that lead to adoption and regular use of SODIS.

Methods

We investigated the behaviours associated with SODIS adoption in households randomly assigned to receive SODIS promotion during a cluster-randomized trial in rural Bolivia. Distinct groups of SODIS-users were identified on the basis of six compliance indicators using principal components and cluster analysis. The probability of adopting SODIS as a function of campaign exposure and household characteristics was evaluated using multinomial models.

Results

Standardised, community-level SODIS-implementation in a rural Bolivian setting was associated with a median SODIS use of 32% (IQR: 17-50). Households that were more likely to use SODIS were those that participated more frequently in SODIS promotional events (OR=1.07, 95%CI: 1.01-1.13), included women (OR=1.18, 95%CI: 1.07-1.30), owned latrines (OR=3.38, 95%CI: 1.07-10.70), and had severely wasted children living in the home (OR=2.17, 95%CI: 1.34-3.49).

Conclusions

Most of the observed household characteristics showed limited potential to predict compliance with a comprehensive, year-long SODIS-promotion campaign reflecting the complexity associated with human behaviour change. However, the findings of this within-group analysis among SODIS-users suggest that the motivation to adopt new water treatment habits and to acquire new knowledge about drinking water treatment is associated with prior engagements in sanitary hygiene and with the experience of contemporary family health concerns.

Household-level factors like the ownership of a latrine, a large proportion of females and the presence of a malnourished child living in a home may be easily assessable indicators for SODIS-programme managers to identify population subgroups that can be targeted for rapid uptake of SODIS.

Background

Systematic reviews of the literature on water, sanitation, and hygiene interventions in developing countries suggest that between 20% and 35% of a total of 3.5 billion diarrhoea episodes per year could be prevented globally by improved drinking water or hand hygiene interventions [1-5]. The evidence to date led the World Health Organisation (WHO) to conclude that household water treatment (HWT) is the most cost-effective approach to reach the United Nations millennium development target 7c of halving the number of persons with no access to safe water (WHO report 2002).

However, the majority of evidence has been collected in controlled intervention studies that document efficacy of HWT by improving water quality and reducing diarrhoeal disease in developing countries [6]. These tightly controlled experiments typically last fewer than six months and include both subsidized (or free) materials and high levels of behaviour reinforcement [7]. Critical issues of effectiveness on a larger scale and sustained use are rarely addressed by these studies [4,8] but are crucial before HWT can be recommended for scaling up [9,10].

Solar water disinfection (SODIS) is one of the simplest and cheapest technologies for household water disinfection. The method relies on disposable translucent plastic bottles of 1-2 litres in which pathogen-containing water is purified by the combined pathogen-inactivating effects of solar radiation and heating [11,12]. Laboratory experiments proved its efficacy in improving the quality of water [12-14]. The method is widely disseminated in developing countries to improve health in settings where safe drinking water is not available. Despite this widespread promotion, only a few field studies assessed its health impact and evidence on acceptance, regular use, and scalability of the method is scarce and inconclusive [9,10,15-18]. Recent studies demonstrate that SODIS promotion is unlikely to reduce diarrhoea in children below 5 years of age if there are low adoption rates and limited long- term use by the target population [6,15,19,20]. It is therefore, important to identify and understand the mechanisms that attenuate the health impacts of SODIS despite its high efficacy for improving water quality under ideal conditions [12,21].

One challenge of assessing the effectiveness of SODIS implementation is the lack of a reliable, unbiased and accepted indicator to measure SODIS-use. Compliance with the SODIS-intervention (e.g. consumption of the SODIS-treated water) is an important

indicator of success of the implementation strategy. To our knowledge, none of the SODIS studies that measured its effectiveness to improve water quality for preventing diarrhoea assessed determinants of compliance directly. To date, the most commonly used end-points to assess SODIS-use rely on self-reported use or the direct observation of water-filled plastic bottles exposed to sunlight [16,18,22-25]. Indicators are often assessed once, usually at the end of the intervention, and the reliability of these indicators is unknown. Self-reported use in response to verbal questioning is known to produce inflated results due to reporting bias [26-29]. Togouet et al. use five measures of self-reported use, direct observation and interviewer opinion to create a 0-5 score to classify 'non-users,' 'irregular users,' and 'regular users' [18]. However, this approach to user classification uses a score that weighs all components equally, and forces the investigator to subjectively choose cut points in that score. There is a need for objective methods to classify households into distinct SODIS user groups.

In this article we present a detailed analysis of SODIS compliance among recipients of a SODIS-intervention who participated in a community-randomised, controlled SODIS trial (cRCT) in rural Bolivia (BoliviaWET). The trial detected no statistically significant reduction in diarrhoea in children under age 5 with an overall SODIS compliance of 32% based on community-health worker assessment [15], a measure that was more conservative than indicators applied in studies with high SODIS-usage rates [16-18]. Here, we use weekly data collected over 12 months from the SODIS compliance monitoring and the SODIS promotion campaign of BoliviaWET to objectively classify households into distinct SODIS-use groups using principal components and cluster analysis. We then use the classified groups to describe the household determinants and campaign implementation factors that are associated with the adoption and utilisation of SODIS in our setting.

Methods

Twenty-two communities from the Totora district, Cochabamba department, Bolivia were included in the cRCT and randomised to receive the SODIS as a HWT. Data of 216 of 225 households enrolled in the 11 intervention communities of the cRCT were included in the analysis. We excluded 9 households from the analysis that were monitored for fewer than 6 weeks over the 12 month follow-up period.

Study site: The Totora district covers an area of 2000 km2. Community settlements are widely dispersed and found at altitudes between 1700 and 3400 metres above sea-level. The majority of the ethnically homogeneous Quechua population are subsistence farmers with small parcels of land growing potatoes, wheat and maize crops. Households keep livestock for their own consumption and for sale. Families typically live in small compounds of three buildings with mud floors, with several persons sleeping in the same room. Only 18% of the homes have a latrine. Most residents defecate in the nearby environment. Unprotected springs are the predominant sources for drinking water.

SODIS campaign: The campaign had two main objectives: i) to create demand for safe drinking water, and ii) to establish a sustainable application of SODIS as a drinking water disinfection method at household level. A locally well-known non-governmental organisation, Project Concern International (PCI), implemented the campaign. PCI has a vast experience in promoting SODIS in rural Bolivian communities. SODIS was introduced during an intensive three-month period before and during the 12-months of field data collection for the trial.

The implementation in intervention communities was standardised at community and household levels. Through participative interactions during district events, community events and home visits, study subjects were introduced to SODIS and environmental health issues related to water and sanitation. District stakeholders from the farmers' union and the official local government, health and school system representatives as well as, formal and informal community leaders were involved in promoting SODIS. In the field, the method was promoted by PCI staff, leaders and advocates, health personnel and teachers, through focus group venues, community- and school events, community training workshops and monthly home visits. Community events were held at least monthly. All community members were invited to these events where they were trained and motivated to practice SODIS daily at their homes.

Experienced health promoters from PCI conducted motivational home visits to empower participants to disinfect their drinking water before consumption and to adopt or improve hygiene habits to create a less contaminated home environment. The motivational home visit strategy was based on participatory hygiene and sanitation transformation methodologies and motivational interviewing [30-32].

SODIS-use assessment: Data regarding SODIS-use were collected by community-based field workers who were integrated into the community and were not involved in any SODIS promotion or implementation activities. Field staff was extensively trained in interviewing and epidemiological observation techniques, data recording, and participatory community motivation approaches. Field staff recorded SODIS-use indicators during weekly home visits with a structured, inconspicuous, observational protocol. In addition, field staff recorded self-reported SODIS-use three months after the beginning and at the end of the intervention campaign (after 15 months).

Table VI.1: Indicators for SODIS-use

Indicator	Rational and Interpretation
SODIS-use indicators	
1. *"Bottles sun-exposed"* Proportion of weeks during which SODIS bottles were observed to be exposed to sunlight (as observed by community-based staff)	Indicator for the intention to disinfect water using SODIS. Indirect indicator to measure actual use.
2. *"Bottles ready-to-drink"* Proportion of weeks during which SODIS bottles were ready-to-drink (as observed by community-based staff)	Households regularely disinfecting water with SODIS usually have bottles of SODIS-treated water ready-to-drink available in-house. Considered to be a more reliable indicator for actual use than "bottles exposed to sunlight"
3. *"Classified user"* Proportion of weeks during which a family was classified as SODIS-user (judgement of community-based staff after observing the family for at least 4 weeks).	Considered as most reliable indicator for actual use. Staff living in the community base their judgement on daily observations of correct application, placing bottles in plain sunlight and/or getting drinking water from a SODIS-bottle when asked for.
4. *"Behavioural change"* Regression coefficient of a logistic regression of the occurrence of bottles exposed to the sunlight (yes/no in a given week) *versus* time.	Indicates behavioural change over time. Coefficient reflects an increase (high values), decrease (low values) or constancy of exposing bottles to sun throughout monitoring time. Note: a coefficient of B=0 indicates constant SODIS-use at high or low levels
Monitoring indicators	
5. *"Time in Study - Bottles sun-exposed"* Total number of weeks during which *"Bottles exposed to sunlight"* was recorded	Discriminates and identifies households with few weeks observed.
6. *"Time in Study - Classified user"* Total number of weeks during which *"Classified user"* was recorded	Discriminates and identifies households with few weeks observed to classify as SODIS-user.

PCI measured study participants' degree of exposure to the SODIS implementation campaign by registering the individual attendance during SODIS promotional events.
In order to arrive at an outcome that describes meaningful types of users, we selected *a priori* four different survey indicators that measure use (Table VI.1). We believe that considering complementary indicators for describing SODIS-use increases the reliability of its measurement by capturing multiple dimensions of potential use. In addition, we use two monitoring indicators (Table VI.1) to identify households that contribute limited information to the classification process due to infrequent observation.

Statistical analysis: To identify patterns of SODIS-use we explored the multivariate distribution of study households in terms of the six quantitative SODIS-use indicators (Table VI.1) by principal component analysis [33]. Identification of meaningful SODIS-user groups was done by Ward's grouping algorithm using R-squared distances as the metric of similarity between households. The Ward's method proved to generate the best classification among several clustering algorithms tested. Five differentiated groups were identified by this approach (Figure VI.1). To confirm the patterns of SODIS-use we further examined the distribution of the study households in the data defined by the factorial axes of a principal component analysis based on the SODIS-use indicators [33].

The effects of the SODIS implementation factors such as the number of times a household member attended a community event, and community- and household level characteristics were tested for univariate differences between groups with the Fisher's exact test for binary data and the Kruskal-Wallis test for non-normally distributed quantitative data. Characteristics with two-sided p-values smaller than 0.1, predictors with less than 25% of missing values to not provoke severe data sparseness problems, and non collinear variables, were retained for inclusion in a multivariable ordinal logistic model. The previously identified SODIS-user groups were used as the categorical-ordinal outcome variable ranging from "non-adopters" to "emerging-adopters". Robust standard errors were calculated to account for community level clustering.

All analyses were performed in STATA 10 (StataCorp. 2007) and in SAS (SAS Institute Inc., Cary, NC, USA).

Ethics: Ethical approval for this study was granted within the framework of the registered BoliviaWET cRCT (ClinicalTrials.gov Identifier: NCT00731497).

Results

Intervention activities and compliance

Household compliance with intervention and morbidity were assessed weekly by the field based monitoring staff for a period of 42 weeks from June 2005 to June 2006 (median: 39 visits, IQR: 34-40).

The SODIS implementation strategy included promotional activities at the community and household level. At the community level, PCI conducted a total of 210 group events, which consisted of 108 community- (median 8 /community, IQR: 7-12), 77 women- (median 7 /community, IQR: 3-10), and 25 school-events (median 3 /community, IQR: 1.5-3). During the study PCI conducted 2886 motivational household visits (median 12 /household, IQR: 8-18).

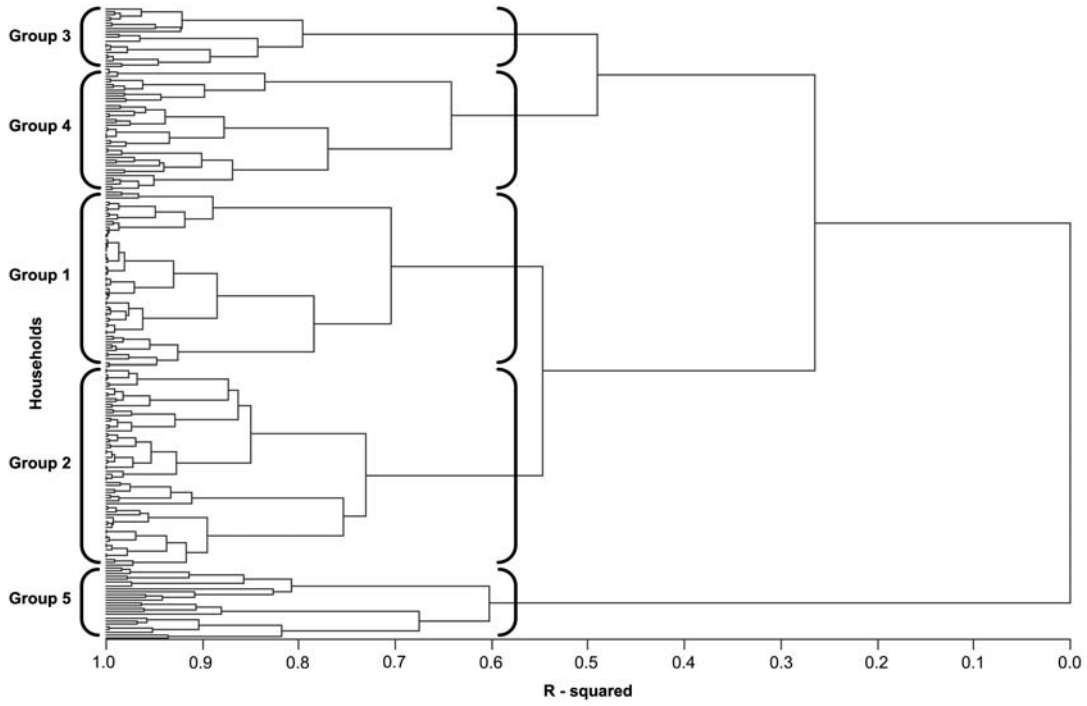
The measured level of SODIS-use varied depending on the indicator used and the source of information. The community-based staff observed an overall median of 33% (IQR: 17-50) of households with SODIS bottles exposed to sunlight during weekly visits. The SODIS-implementing PCI staff registered during monthly household visits a median proportion of 75% (IQR: 60-85) of households with SODIS bottles exposed to the sun. After three months of intensive implementation, PCI staff recorded 77% of household respondents reporting regular SODIS-use, and 88% at the end of the study.

SODIS-user group classification

Figure VI.1 summarizes the results of the cluster analysis, which identified five distinct SODIS-use groups based on household-level use indicators: Group 1 = ‘non-adopters’, Group 2 = ‘minimal-adopters’, Group 3 = ‘declining-adopters’ and group 4 = ‘emerging-adopters’ (see also supplementary Figure VI.S1). Groups 3 and 4 comprised households with the highest SODIS-usage rates; group 3 with an initially high uptake and declining SODIS-use over time, group 4 with an emerging adoption pattern. Based on this group separation, we used characteristics of households in the groups to describe them in meaningful, qualitative terms. Figure VI.2 shows the difference between groups in four different SODIS-use indicators (self-reported and observed use) and two monitoring indicators (Table VI.1), and Figure VI.3 shows different SODIS-usage rates over time using the same indicators for the five user groups. Group 5 (25 households) differed from the other groups with respect to the time under observation (indicators 4 and 5): Its time

under observation (median 20 weeks, IQR: 16-23) was considered too short to obtain a valid estimate of SODIS-use and led to high variability in all of the indicators (Fig 2e). Based on the limited information in group 5, we decided to exclude it from between-group comparisons in the ordinal logistic model.

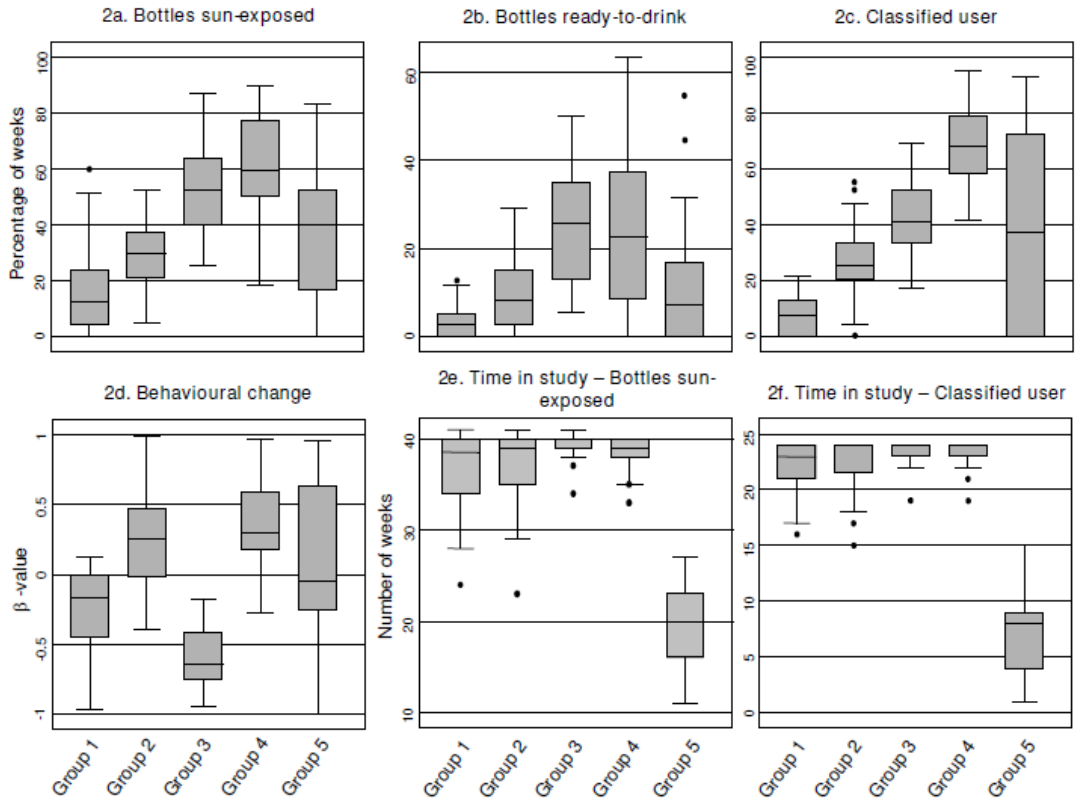
Figure VI.1: Dendogram with the grouping history of the hierarchical classification (Ward's method).



Legend: Horizontal axis denotes the linkage distance (R-square distance) between households according to their SODIS-use indicators listed in Table VI.1

The group of 'non-adopters' consisted of households with little interest in adopting and using SODIS (median proportion of weeks with bottles exposed to sun were observed: 0.13; IQR: 0.04-0.24) (Fig. 2a and 3a). 'Minimal-adopters' used SODIS more frequently: median proportion: 0.3 (IQR: 0.21-0.38) (Fig. 2a and 3b) of the weeks observed. The 'declining- and emerging adopters' constituted the households with the highest SODIS-usage rates (median: 0.53 and 0.60; IQR: 0.40-0.64 and 0.50-0.78) (Fig. 2a and 3c and 3d). 'Declining-adopters' used SODIS more often at the beginning of the follow-up (Indicator 4 "Behavioral change" in Table VI.1, logistic regression coefficient bottles exposed to sun vs. time) median: -0.65; IQR: -0.75-0.38 (Fig. 2d and 3c). 'Emerging-adopters' used SODIS more often toward the end of the follow-up with a median of 0.30; IQR: 0.20-0.60 (Fig. 2d and 3d).

Figure VI.2: Box-plots of 5 SODIS-user groups differing in 6 SODIS-use indicators (see Table VI.1)

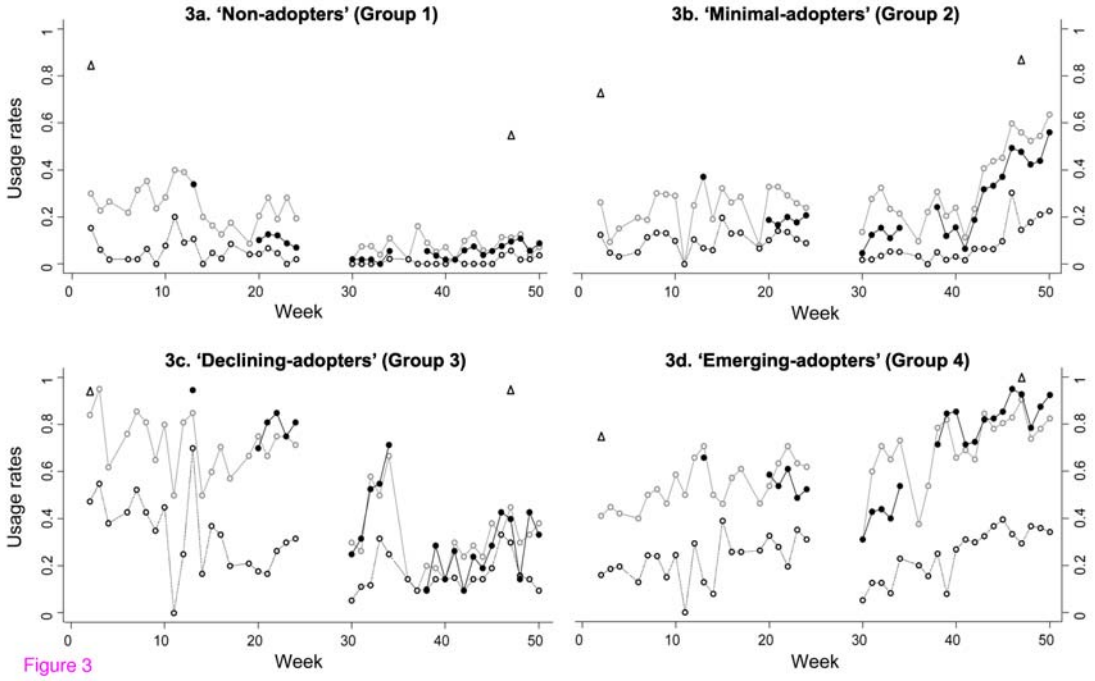


Factors influencing SODIS adoption

The characteristics of the different SODIS user groups comparing in a univariate analysis ‘non-adopters’, ‘minimal-adopters’, and the two frequent user groups of ‘declining-‘ and ‘emerging-adopters’ are presented in Table VI.2.

Some household characteristics differed significantly at a 95%-confidence level between SODIS-use groups. Households with the highest SODIS-usage rates exhibited the following specific features: ‘Emerging-adopters’ consisted of more females compared to the other groups. ’Decreasing-adopters’ were more likely to own bicycles. Households from both of the higher user-groups were more likely to own a latrine (56% and 26%) than ‘non- and minimal- adopters’ households (both 8%). Further, they were more likely to have severely wasted children (two times substandard weight-for-height = 65% and 66%, respectively) than ’non-adopters’ (17%) and ‘minimal-adopters’ (25%). ‘Non-adopters’ lived the furthest distance away from their water source with a median of 100m, followed by the ‘minimal-adopters’ (30m). In contrast, distances to the water source were much shorter for households with the highest SODIS-usage rates (5m and 10m in ’declining-and emerging-adopters’).

Figure VI.3: Weekly observed proportion of households using SODIS in five SODIS-user groups



Legend: Legend: Open triangles: self-reported SODIS-use at the beginning (after 3 month of initial SODIS promotion) and at the end of follow-up; filled dots: SODIS-use observed by project staff living in the community (see Table VI.1 for definition); open grey circles: SODIS bottles observed on the roof; open black circles: SODIS bottles observed ready to drink

Table VI.2. Distribution of potential household determinants of SODIS-use

				Groups based on SODIS-use behaviour												
	Total			Group 1 ('non-adopters')			Group 2 ('minimal-adopters')			Group 3 ('declining-adopters')			Group 4 ('emerging-adopters')			p-values*
	n=216			n=60			n=68			n=21			n=42			
Demography																
Number of household members	216	6.0	(5;9)	60	6.0	(5;8.5)	68	6.0	(4;9)	21	6.0	(5;7)	42	7.0	(5;9)	0.24
Age of household members	216	15.8	(13;18.1)	60	15.5	(13.7;17.6)	68	15.9	(13.3;18.7)	21	15.9	(13;17.8)	42	16.0	(12.1;18.4)	0.88
Number of females	216	3.0	(2;4)	60	3.0	(2;4)	68	3.0	(2;5)	21	3.0	(2;4)	42	4.0	(3;6)	**0.04**
Pregnant women at start of campaign	216	0.0	(0;0)	60	0.0	(0;0)	68	0.0	(0;0)	21	0.0	(0;0)	42	0.0	(0;0)	**0.09**
Children aged < 5	216	1.0	(1;2)	60	1.0	(1;2)	68	2.0	(1;2)	21	1.0	(1;2)	42	2.0	(1;2)	**0.06**
Children aged 5-9	216	1.0	(0;2)	60	2.0	(0;2)	68	1.0	(0;2)	21	1.0	(0;2)	42	1.0	(1;2)	0.60
Children aged 10-14	216	0.0	(0;1)	60	0.0	(0;2)	68	0.0	(0;1)	21	0.0	(0;1)	42	0.0	(0;2)	0.80
Members aged 15-19	216	0.0	(0;1)	60	0.0	(0;1)	68	1.0	(0;1)	21	0.0	(0;1)	42	0.0	(0;1)	0.95
Members aged > = 20	216	2.0	(2;2)	60	2.0	(2;2)	68	2.0	(2;2.5)	21	2.0	(2;2)	42	2.0	(2;2)	0.17
Caregivers' age	208	28.0	(23;36)	58	28.8	(23;35)	67	29.0	(23;37)	19	30.0	(22;36)	41	28.0	(23;40)	0.87
Socioeconomic characteristics																
Years of household heads' schooling	155	4.0	(3;5)	43	4.0	(2;5)	52	4.0	(3;5)	14	4.0	(3;5)	28	4.0	(2.5;5)	0.38
Monthly household income in US$	120	16.9	(0;37.5)	35	12.5	(0;25)	37	12.5	(0;31.3)	7	25.0	(12.5;37.5)	24	31.3	(0;47.5)	0.25
Bicycle: n (%)	192	107	(55.7)	49	25	(51)	65	40	(61.5)	18	14.0	(77.8)	39	17.0	(43.6)	**0.07**
Radio: n (%)	192	158	(82.3)	49	41	(83.7)	65	55	(84.6)	18	13	(72.2)	39	30	(76.9)	0.53
Gas cooker: n (%)	181	32	(17.7)	53	9	(17)	66	16	(24.2)	20	3	(15)	42	4	(9.5)	0.28
Number of rooms	192	3.0	(2;4)	49	2.0	(2;3)	65	3.0	(2;4)	18	3.0	(2;3)	39	3.0	(2;4)	0.29
Latrine: n (%)	192	34	(17.7)	49	4	(8.2)	65	5	(7.7)	18	10	(55.6)	39	10	(25.6)	**>0.001**
Electricity: n (%)	192	36	(18.8)	49	11	(22.5)	65	16	(24.6)	18	1	(5.6)	39	3	(7.7)	**0.06**
Solar panel: n (%)	130	30	(23)	29	9	(31)	50	8	(16)	12	3	(25)	25	5	(20)	0.44
Tiled roof: n (%)	181	57	(31.5)	53	19	(35.9)	66	18	(27.3)	20	8	(40)	42	12	(28.6)	0.60
Environmental housing factors																
Use of improved water source: n (%)**	192	149	(77.6)	49	36	(73.5)	65	53	(81.5)	18	15	(83.3)	39	29	(74.4)	0.69
Use of unimproved water source: n (%)***	192	133	(69.3)	49	37	(75.5)	65	48	(73.9)	18	10	(55.6)	39	24	(61.5)	0.23
Distance to water source in metres.	192	22.5	(5;50)	49	100.0	(10;200)	65	30.0	(7;100)	18	5.0	(4;30)	39	10.0	(5;200)	**0.03**
Turbidity of source water (NTU)	101	5.0	(5;20)	30	5.0	(5;40)	34	5.0	(5;20)	7	5.0	(5;40)	19	5.0	(5;15)	0.79
Faecal contamination of housing environment: n (%)	185	106	(57.3)	50	31	(62)	62	34	(54.8)	16	9	(56.3)	36	16	(44.4)	0.46
Animals present in the kitchen: n (%)	168	45	(26.8)	42	15	(35.7)	57	13	(22.8)	15	3	(20)	33	6	(18.2)	0.32
Soap, detergent present in the kitchen: n (%)	166	29	(17.5)	41	6	(14.6)	56	12	(21.4)	15	3	(20)	33	4	(12.1)	0.68
Household members health status																

Households with at least one stunted child < 5: n (%)	167	62	(37.1)	43	12	(27.9)	53	19	(35.9)	17	8	(47.1)	35	16	(45.7)	0.33
Households with at least one wasted child < 5: n (%)	167	85	(50.9)	43	17	(39.5)	53	25	(47.2)	17	11	(64.7)	35	23	(65.7)	**0.08**
Diarrhoea incidence in children < 5 before start of intervention	216	3.0	(0;7)	60	3.0	(0;5)	68	3.0	(1;7)	21	6.0	(1;12)	42	3.0	(0;6)	0.22
Diarrhoea prevalence (%) in children < 5 before start of intervention	216	7.0	(1;14)	60	5.0	(0;12)	68	8.0	(2;14)	21	7.0	(2;32)	42	6.0	(0;17)	0.26
Cough prevalence (%) in children < 5 before start of intervention	216	8.0	(0;20)	60	5.0	(0;17)	68	8.0	(2;21)	21	1.0	(0;17)	42	10.0	(0;24)	0.72
Fever prevalence (%) in children < 5 before start of intervention	216	7.0	(2;15)	60	6.0	(0;16)	68	5.0	(1;11)	21	6.0	(2;19)	42	7.0	(2;22)	0.58
Hand-washing behaviour																
Hand-washing per day of children > 5 and adults	169	4.0	(3;5)	44	4.0	(3;5)	57	3.0	(3;5)	15	3.0	(3;5)	33	4.0	(3;5)	0.27
Hand-washing per day of children < 5	192	2.6	(2;3)	49	2.5	(2;3)	65	2.5	(2;3)	18	3.0	(2;3)	39	2.7	(2;3)	0.96
Household water management																
Safe storage: n (%)	155	19	(12.3)	34	4	(11.8)	57	5	(8.8)	14	5	(37.5)	30	2	(6.7)	**0.06**
Water disinfection: n (%)	192	42	(21.9)	49	12	(4.5)	65	13	(20)	18	5	(27.8)	39	9	(23.1)	0.86
Household water consumption [l / household day]	189	40	(20;50)	58	35	(20;50)	67	40	(20;60)	21	50	(20;60)	41	30	(20;60)	0.81
Satisfied with quality of drinking water: n (%)	201	190	(94.5)	54	51	(94.4)	64	59	(92.2)	19	19	(94.7)	40	39	(97.5)	0.76

Legend: Baseline data are median (Q1; Q3), otherwise specified. *: Kruskal-Wallis and Fisher's exact test for comparing group 1, 2, 3, and 4; **: Improved water source: piped water into dwelling, plot or yard; tubewell/borehole; protected spring; rainwater collection. ***: Unimproved water source: unprotected dug well or spring; bowser-truck; surface water (river, dam, pond, irrigation channels)

Table VI.3 summarizes household exposure to the SODIS campaign through active participation at community-level events and through passive exposure to motivational activities during household visits. Since the implementation was standardised at community- and household levels there is no difference between the four SODIS-user groups regarding campaign features such as ‘Number of events taken place per community’, ‘Average number of participants per event and community’, and ‘Number of household visits per household’. However, groups differed significantly regarding active participation at those events. ‘Non-adopters’ participated on average at half of the events offered, whereas ‘declining and emerging adopters’ participated at 78% and 71% of the events. The level of participation at school events was similar across groups, since participation was mandatory for school children in all schools in the study site.

Since SODIS implementation indicators were correlated with each other, only one indicator (‘Total number of events visited by at least one household member’) was included in the model because it encapsulates the others. Table VI.4 presents results of the ordinal logistic regression model. The model containing only the SODIS implementation factor revealed that ‘Total number of events visited by at least one household member’ is positively associated with frequent SODIS use group membership. For each additional event visited the odds of being in the next higher category of adoption was 1.07 (95% CI : 1.01-1.13). The multivariable model showed that higher adoption groups were more likely to own a latrine (OR: 3.38; 95% CI: 1.07-10.70) and to have at least one wasted child living in the household (OR: 2.17; 95% CI: 1.34-3.49). Furthermore, the number of females living in a household was significantly associated with group membership prediction (OR: 1.18; 95% CI: 1.07-1.30).

Table VI.3. SODIS campaign at household and community level

	Total n=216			Groups based on SODIS-use behaviour: Group 1 ('non-adopters') n=60			Group 2 ('minimal-adopters') n=68			Group 3 ('declining-adopters') n=21			Group 4 ('emerging-adopters') n=42			p-values *
Household exposure to SODIS campaign																
Different events visited by at least one household member (n)	213	10.0	(6;13)	58	7.5	(6;12)	68	10.0	(6;12)	21	13.0	(9;17)	42	12.0	(7;14)	**0.002**
Events visited by at least one household member (n)	213	11.0	(6;15)	58	8.5	(6;14)	68	11.0	(6;15)	21	16.0	(11;22)	42	14.0	(10;18)	**0.004**
Proportion of possible events per community visited (%)	213	62.0	(39;83)	58	50.0	(32;80)	68	62.0	(44;81)	21	78.0	(57;100)	42	71.0	(48;94)	**0.017**
Events visited by most active household member (n)	213	6.0	(4;8)	58	5.0	(3;8)	68	5.5	(4;8)	21	9.0	(6;11)	42	6.0	(4;9)	**0.002**
Community events visited by at least one household member (n)	213	5.0	(3;7)	58	5.0	(3;7)	68	5.0	(3;9)	21	5.0	(3;6)	42	7.0	(5;9)	**0.019**
Women events visited by at least one household member (n)	213	2.0	(1;4)	58	2.0	(1;3)	68	2.0	(1;4)	21	7.0	(2;8)	42	3.0	(1;4)	**0.003**
School events visited by at least one household member (n)	213	0.0	(0;2)	58	0.0	(0;2)	68	0.0	(0;2)	21	0.0	(0;3)	42	0.0	(0;3)	0.515
Household visits by promoting NGO (n)	213	12.0	(8;18)	57	10.0	(6;19)	68	13.0	(9;18)	21	16.0	(12;21)	42	12.5	(9;18)	0.224
SODIS campaign at community level																
Events taken place per community (n)	216	18.0	(16;21)	60	19.0	(15.5;21)	68	18.0	(16.5;21)	21	21.0	(17;23)	42	17.5	(16;21)	**0.037**
Average number of participants per event per community	216	29.6	(23.2;40.4)	60	29.4	(20.1;40.4)	68	30.1	(24.0;48.8)	21	27.1	(27.1;30.1)	42	30.1	(27.1;40.4)	**0.071**
Average duration of events per community (hrs)	216	3.3	(2.9;3.8)	60	3.1	(2.8;3.6)	68	3.2	(2.8;3.7)	21	3.8	(3.4;3.8)	42	3.4	(3.1;3.8)	**0.018**

Data are median (Q1;Q3), otherwise specified. *: Kruskal-Wallis and Fisher's exact

Discussion

We characterised in a cluster analysis five distinct SODIS user groups after a 15-month comprehensive SODIS-dissemination campaign among the participants of a community-randomised, controlled SODIS-evaluation trial in rural Bolivia.

Household characteristics that were most strongly associated with the adoption of the SODIS household water treatment method include the intensity of exposure to the SODIS campaign, the number of females per household, latrine ownership, and having severely wasted children living in the home. The knowledge of household factors found to be related to SODIS-use may help to target populations that would more easily adopt SODIS and, therefore, benefit most from SODIS implementations.

Table VI.4. Results of the ordinal logistic regression models

Predictor	**Univariable model (n=189) (SODIS implementation factor only)**		
	OR	**95% CI***	**P value**
Total no. of events visited by at least one household member	1.07	1.01-1.13	**0.02**
	Multivariable model (n = 146)		
	OR	**95% CI***	**P value**
Total no. of events visited by at least one household member	1.04	0.98-1.11	0.15
Nr of females per household	1.18	1.07-1.30	**0.001**
Household with pregnant women at start of campaign	1.33	0.67-2.64	0.41
Bicycle owenrship	0.75	0.35-1.64	0.48
Latrine	3.38	1.07-10.70	**0.04**
Distance to water source (log of)	0.94	0.73-1.22	0.65
Households with at least one wasted child under 5	2.17	1.34-3.49	**0.001**

* calculated from robust standard errors adjusted for community cluster

Our findings suggest that the motivation to adopt new water treatment habits and to acquire new knowledge about drinking water treatment is associated with prior health-related engagements, e.g. in latrine construction, and by with the experience of family health concerns such as living with an acutely malnourished child. In addition, higher SODIS-use was associated with the frequency of exposure to SODIS promotion of anyone of the household members. It is likely that eager adopters of new ideas and technological inventions such as SODIS are more interested in participating at the related promotional events.

Our findings are consistent with previous studies: In a similar setting in Bolivia, Moser and Mosler [25] found existing knowledge about the need to treat drinking water predicted early SODIS adoption. Applying the theory of the diffusion of innovations from Rogers et al. [34] in a SODIS diffusion programme in rural Bolivia they found that participation at SODIS-campaign events correlated positively with SODIS-use [24]. Further, a field study from Nicaragua reported that intention to use and actual use were related to a positive attitude toward the new technology [35]. These coherent findings on the motivating factors for SODIS adoption underscore the importance of determining a target population's characteristics and its attitude towards new technology prior to promoting SODIS.

The indicators we employed in our analysis to measure households' weekly SODIS-use were based on inconspicuous structured observations conducted by our community-based staff who were not involved in any SODIS-promotion activity. In combining objective indicators measuring, visible signs of use (e.g. bottles exposed to sun) with proxies more responsive to the direction and magnitude of the change of treatment behaviour (e.g. weekly observation of correct application of SODIS) we increased the quality of measurement and reduced the potential for reporting bias and misclassification error [26-28]. Our independent evaluation of SODIS-use generated much lower adoption rates than estimates from the implementing organization, PCI (32% versus 75%). This underscores the potential for bias in situations when implementers evaluate their own work. Such courtesy bias and over-reporting of compliance with the intervention is well known from water, sanitation and hygiene intervention studies [7,26,36-42]. The discrepancy between the levels of SODIS compliance assessed through different indicators in our study raises questions about the interpretation of compliance rates of both, studies in peer-reviewed and grey literature. Our results highlight the importance of choosing independent staff and a valid and responsive indicator to assess use and to draw conclusions about the implementation effectiveness of HWT intervention programmes.

Despite an intensive 15-month promotion campaign carried out by a highly qualified implementing organization, we observed 32% overall compliance with the solar water disinfection method during our 12 months of follow-up [15]. Our findings suggest that SODIS promotion would benefit from re-assessing the core marketing messages and approaches to reach the critical 50% fraction of early and willing SODIS adopters in the population [25]. Our analysis identified some characteristics associated with frequent use.

However, it is the characteristics of willing but occasional user groups (our ‘minimal adopters’) to whom new marketing and promotion strategies should be targeted [43]. However, based on the characteristics that we measured, it was difficult to differentiate the ‘minimal adopters’ from ‘non-adopters’ (Table VI.2). In this population, the ‘non-adopter’ and ‘minimal-adopter’ groups included the most marginalized households by observable characteristics: they were poorer, lived further from water sources, rarely owned a latrine, had more frequently faecally contaminated home environments, and had more animals roaming their kitchen area; yet, unexpectedly, they were less likely to have stunted or wasted children in their families (Table VI.2).

Criteria to plan for the successful roll-out and targeting of water and sanitation programmes have often been suggested [44]. In the Bolivian context SODIS-programme planning may benefit from assessing easy measurable household-level factors like the ownership of a latrine, a large proportion of females and the presence of a malnourished child to identify population subgroups that can be targeted for rapid uptake of the SODIS HWT method.

There are limitations to this study. The participating communities were not homogenous regarding pre-existing water supplies and sanitation infrastructures, previous exposure to sanitation and hygiene campaigns, as well as political support to participate in the study. Further, the ordinal logistic regression assumes that the categories follow an intrinsic order. This order is evident for ‘non- and minimal adopters’ but is less obvious in the case of ‘declining- and emerging-adopters’. However, from the programme-implementation viewpoint the sustained user, i.e. the ‘emerging adopters’, are, of course, the most important group. To ensure that our findings were not sensitive to the modeling approach, we repeated the analysis using multinomial regression, which does not impose an order to the categorical outcome. Analogous to our presented results, the multinomial regression identified latrine ownership and presence of severely wasted children as the most important predictors of SODIS-use categories (data not shown). Finally, data on the SODIS-use indicator ‘Households rated as SODIS-user by implementation-independent field worker’, was incomplete because (i) the indicator was implemented after an intensive 3-month pilot phase, and (ii) it required the randomly-rotated field staff (every 3 months) to familiarize themselves with each local community for a period of four weeks before they could report the indicator [15]. While we believe this measure reduced

systematic reporting bias and enhanced the reliability of SODIS-use measurement, it reduced the total observation time available for analysis.

Conclusions

Analyses of implementation effectiveness and the dynamics of SODIS-uptake from large-scale SODIS dissemination programmes are rarely published. Our findings suggest that households that have more women, own a latrine, have malnourished (wasted) children and are close to their water source are more likely to adopt SODIS during an intensive promotion campaign. Households that did not adopt SODIS tend to be poorer, further from water sources and have less hygienic home environments. This finding suggests how implementers could identify populations most likely to use (initially and over a sustained period) and benefit from SODIS interventions.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AC and DM conceived the idea and developed the design for the study. AC wrote the original draft manuscript, and incorporated revisions from each of the co-authors. GDP and JH contributed to the conception and design of the manuscript and conducted the statistical analysis. AC and MC coordinated and supervised data acquisition. DM, JH, GDP, and BFA wrote parts of the paper and together with, MC, JMC, and SI contributed to the conception of the manuscript and provided revisions. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge the communities and the authorities of the Totora District and the SODIS implementation team around Carlos Morante and Luciano Cespedes for their relentless efforts to intertwine research and development approaches. We acknowledge the collaboration with Project Concern International, Cochabamba who thankfully accommodated and actively engaged in this research work as part of their routine rural programme implementation. We specifically thank Fidel Alvarez, former director of

Project Concern International, Cochabamba for his institutional and personal support. We are grateful to the study team: Roy Cordova, Freddy Arauco, Abrahan Cuevas, Fernando Salvatierra, and the entire MMS (Morbidity Monitoring Staff). Michael Hobbins and Monica Daigl provided valuable inputs to the study implementation.

Funding for the BoliviaWET trial was provided by the National Institutes of Health, Award Number R01AI50087-01. In kind and financial support were also provided by Project Concern International, University of San Simon, SODIS Foundation, Cochabamba and the Swiss Tropical and Public Health Institute, Basel.

The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; the preparation, review, or approval of the manuscript.

References

1. Aiello AE, Coulborn RM, Perez V, Larson EL: **Effect of hand hygiene on infectious disease risk in the community setting: a meta-analysis.** *Am J Public Health* 2008, **98:**1372-1381.

2. Arnold BF, Colford JM, Jr.: **Treating water with chlorine at point-of-use to improve water quality and reduce child diarrhea in developing countries: a systematic review and meta-analysis.** *Am J Trop Med Hyg* 2007, **76:**354-364.

3. Clasen T, Roberts I, Rabie T, Schmidt W, Cairncross S: **Intervention to improve water quality for preventing diarrhoea.** *Cochrane Database of Systematic Reviews* 2006.

4. Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, Colford JM, Jr.: **Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis.** *Lancet Infect Dis* 2005, **5:**42-52.

5. Waddington H, Snilstveit B: **Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea.** *J Develop Effectiveness* 2009, **1:**295-335.

6. Sobsey MD, Stauber CE, Casanova LM, Brown JM, Elliott MA: **Point of use household drinking water filtration: A practical, effective solution for providing sustained access to safe drinking water in the developing world.** *Environ Sci Technol* 2008, **42:**4261-4267.

7. Luby SP, Mendoza C, Keswick BH, Chiller TM, Hoekstra RM: **Difficulties in bringing point-of-use water treatment to scale in rural Guatemala.** *Am J Trop Med Hyg* 2008, **78:**382-387.

8. Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S: **Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis.** *BMJ* 2007, **334:**782.

9. Hunter PR: **Household water treatment in developing countries: comparing different intervention types using meta-regression.** *Environ Sci Technol* 2009, **43:**8991-8997.

10. Schmidt WP, Cairncross S: **Household water treatment in poor populations: is there enough evidence for scaling up now?** *Environ Sci Technol* 2009, **43:**986-992.

11. McGuigan KG, Joyce TM, Conroy RM, Gillespie JB, Elmore-Meegan M: **Solar disinfection of drinking water contained in transparent plastic bottles: characterizing the bacterial inactivation process.** *J Appl Microbiol* 1998, **84:**1138-1148.

12. Wegelin M, Canonica S, Mechsner K, Fleischmann T, Pesario F, Metzler A: **Solar water disinfection (SODIS): Scope of the process and analysis of radiation experiments.** *J Water SRT-Aqua* 1994, **43:**154-169.

13. Joyce T, Kenny V, McGuigan K, Barnes J: **Disinfection of water by sunlight.** *Lancet* 1992, **340:**921.

14. Boyle M, Sichel C, Fernandez-Ibanez P, rias-Quiroz GB, Iriarte-Puna M, Mercado A, Ubomba-Jaswa E, McGuigan KG: **Bactericidal effect of solar water disinfection under real sunlight conditions.** *Appl Environ Microbiol* 2008, **74:**2997-3001.

15. Mäusezahl D, Christen A, Pacheco GD, Tellez FA, Iriarte M, Zapata ME, Cevallos M, Hattendorf J, Cattaneo MD, Arnold B et al.: **Solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia: a cluster-randomized, controlled trial.** *PLoS Med* 2009, **6:**e1000125.

16. Rose A, Roy S, Abraham V, Holmgren G, George K, Balraj V, Abraham S, Muliyil J, Joseph A, Kang G: **Solar disinfection of water for diarrhoeal prevention in southern India.** *Arch Dis Child* 2006, **91:**139-141.

17. Conroy RM, Elmore-Meegan M, Joyce T, McGuigan KG, Barnes J: **Solar disinfection of drinking water and diarrhoea in Maasai children: a controlled field trial.** *Lancet* 1996, **348:**1695-1697.

18. Togouet SZ, Graf J, Gangoue Pieboji J, Kemka N, Niyitegeka D, Meierhofer R. **Health gains from solar water disinfection (SODIS): evaluation of a water quality intervention in Yaoundé, Cameroon.** *J Water Health* 2010, In Press, Uncorrected Proof. doi:10.2166/wh.2010.003

19. Arnold B, Arana B, Mausezahl D, Hubbard A, Colford JM, Jr.: **Evaluation of a pre-existing, 3-year household water treatment and handwashing intervention in rural Guatemala.** *Int J Epidemiol* 2009, **38:**1651-1661.

20. Rainey RC, Harding AK: **Acceptability of solar disinfection of drinking water treatment in Kathmandu Valley, Nepal.** *Int J Environ Health Res* 2005, **15:**361-372.

21. Sommer B, Marino A, Solarte Y, Salas ML, Dierolf C, aliente C, ora D, echsteiner R, etter P, irojanagud W et al.: **SODIS - an emerging water treatment process.** *J Water SRT-Aqua* 1997, **46:**127-137.

22. Kraemer SM, Mosler HJ: **Persuasion factors influencing the decision to use sustainable household water treatment.** *Int J Environ Health Res* 2010, **20:**61-79.

23. Graf J, Meierhofer R, Wegelin M, Mosler HJ: **Water disinfection and hygiene behaviour in an urban slum in Kenya: impact on childhood diarrhoea and influence of beliefs.** *Int J Environ Health Res* 2008, **18:**335-355.

24. Heri S, Mosler HJ: **Factors affecting the diffusion of solar water disinfection: a field study in Bolivia.** *Health Educ Behav* 2008, **35:**541-560.

25. Moser S, Mosler HJ: **Differences in influence patterns between groups predicting the adoption of a solar disinfection technology for drinking water in Bolivia.** *Soc Sci Med* 2008, **67:**497-504.

26. Biran A, Rabie T, Schmidt W, Juvekar S, Hirve S, Curtis V: **Comparing the performance of indicators of hand-washing practices in rural Indian households.** *Trop Med Int Health* 2008, **13:**278-285.

27. Cousens S, Kanki B, Toure S, Diallo I, Curtis V: **Reactivity and repeatability of hygiene behaviour: structured observations from Burkina Faso.** *Soc Sci Med* 1996, **43:**1299-1308.

28. Curtis V, Cousens S, Mertens T, Traore E, Kanki B, Diallo I: **Structured observations of hygiene behaviours in Burkina Faso: validity, variability, and utility.** *Bull World Health Organ* 1993, **71:**23-32.

29. Gittelsohn J, Shankar AV, West KP, Ram RM, Gnywali T: **Estimating Reactivity in Direct Observation Studies of Health Behaviors.** *Human Organization* 1997, **56:**182-189.

30. Narayan D: *Participatory Evaluation.* Washington,D.C.: The World Bank; 1993.

31. Srinivasan L: *Tools for Community Participation: A Manual for training trainers in participatory techniques.* New York: United Nations Development Programme; 1990.

32. World Health Organisation: *PHAST step-by-step guide: A participatory approach for the control of diarrhoeal diseases.* Geneva, Switzerland: World Health Organization; 1998.

33. Lebart L, Piron M, Morineau A: *Statistique exploratoire multidimensionale.* Paris: Dunod; 2000.

34. Rogers EM: *Diffusion of innovations.* New York: Free Press; 2003.

35. Altherr AM, Mosler HJ, Tobias R, Butera F: **Attitudinal and Relational Factors Predicting the Use of Solar Water Disinfection: A Field Study in Nicaragua.** *Health Educ Behav* 2008, **35:**207-220.

36. Almedom AM, Blumenthal U, Manderson L: *Hygiene Evaluation Procedures.* Boston, MA, USA: International Nutrition Foundation for Developing Countries; 1997.

37. Chiller TM, Mendoza CE, Lopez MB, Alvarez M, Hoekstra RM, Keswick BH, Luby SP: **Reducing diarrhoea in Guatemalan children: randomized controlled trial of flocculant-disinfectant for drinking-water.** *Bull World Health Organ* 2006, **84:**28-35.

38. Gupta SK, Islam MS, Johnston R, Ram PK, Luby SP: **The chulli water purifier: acceptability and effectiveness of an innovative strategy for household water treatment in Bangladesh.** *Am J Trop Med Hyg* 2008, **78:**979-984.

39. Luby SP, Agboatwalla M, Feikin DR, Painter J, Billhimer W, Altaf A, Hoekstra RM: **Effect of handwashing on child health: a randomised controlled trial.** *Lancet* 2005, **366:**225-233.

40. Luby SP, Agboatwalla M, Painter J, Altaf A, Billhimer W, Keswick B, Hoekstra RM: **Combining drinking water treatment and hand washing for**

diarrhoea prevention, a cluster randomised controlled trial. *Trop Med Int Health* 2006, **11:**479-489.

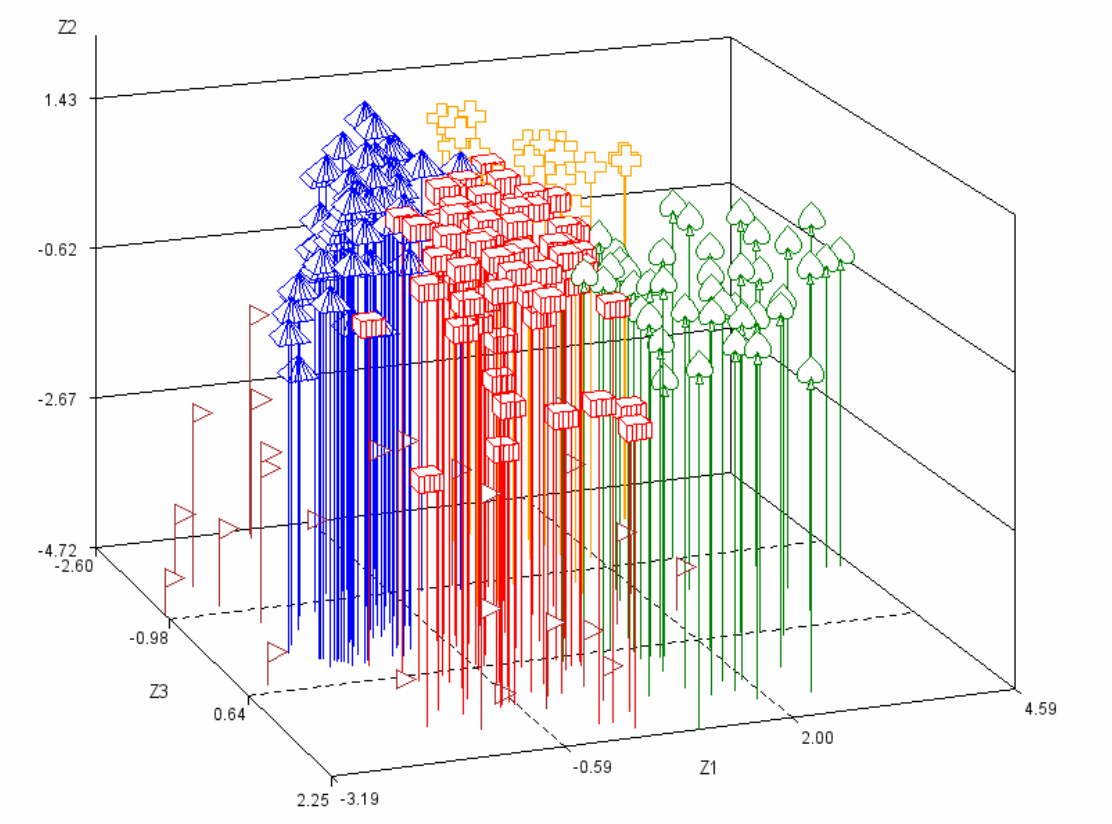
41. Luby SP, Agboatwalla M, Painter J, Altaf A, Billhimer WL, Hoekstra RM: **Effect of intensive handwashing promotion on childhood diarrhea in high-risk communities in Pakistan: a randomized controlled trial.** *JAMA* 2004, **291:**2547-2554.

42. Sandora TJ, Taveras EM, Shih MC, Resnick EA, Lee GM, Ross-Degnan D, Goldmann DA: **A randomized, controlled trial of a multifaceted intervention including alcohol-based hand sanitizer and hand-hygiene education to reduce illness transmission in the home.** *Pediatrics* 2005, **116:**587-594.

43. Tamas A, Tobias R, Mosler HJ: **Promotion of solar water disinfection: comparing the effectiveness of different strategies in a longitudinal field study in Bolivia.** *Health Commun* 2009, **24:**711-722.

44. Samanta BB, Van Wijk CA: **Criteria for successful sanitation programmes in low income countries.** *Health Policy Plan* 1998, **13:**78-86.

Figure VI.S1. 3D scatter plot view of SODIS user groups of the first three principal components.



Chapter VII:

Discussion and concluding remarks

1. Discussion and Concluding Remarks.

The BoliviaWET experience of analysing the effect of solar water disinfection on childhood diarrhoea provided abundant material for statistical research. The main analysis of the trial (chapter II) motivated the assessment of analytical methods for cluster randomized trials, under situations similar to BolivaWET data, i.e. overdispersed count data, variation of individual follow-up periods, cluster size imbalance, levels of clustering, sample size (chapter III). We also compared the performance of methods for point and interval estimation of a clustering measure in similar situations (chapter IV). We evaluated the local term "*k'echalera*", in the Quechua language, as a means to assess the diarrhoeal syndrome (chapter V). Finally, we explored the meaning of SODIS-use from a multivariate perspective, identified typologies of SODIS-users and identified the factors that influence on the adoption of SODIS (chapter VI).

This material was originally conceived as a set of instruments to validate the primary and secondary analyses of the trial. Additionally, it provided elements to enrich the interpretation of the trial results. We consider, however, that this work is relevant to community randomized trials in general and to home-based water treatment interventions to prevent diarrhoea in particular.

The next section of this discussion considers our main findings in context of the design of new cluster-unit trials. A further section considers the implications for methods of analysis of the results. This is followed by a section that focuses on the implications of overdispersion. Next some more general remarks on the statistical methods applied in chapters V and VI are presented. Finally, the overall conclusions of this thesis are provided.

Design aspects

Pair-matching is particularly recommended in community randomized trials because disparity between trial arms is more likely if the total number of clusters is limited [1].

Matching clusters prior to randomization by factors related to the outcome can thus make randomization much more effective especially if the clusters are heterogeneous [2]. The BoliviaWET trial considered pair-matching to reduce the chance of assigning the treatment to inherently different communities in terms of diarrhoea rates. It was also assumeed that controlling the outcome would indirectly assure balanced risk and confounding factors at baseline between arms [2]. Consequently, communities were matched into pairs by baseline diarrhoea incidence. As observed in chapter I, the between-pairs variance was estimated to be zero, reflecting a lack of control in terms of the outcome variation (Tables I.3, III.3). However, other baseline characteristics were fairly well balanced between the study arms (Table II.1). A few exceptions were some water management and consumption characteristics.

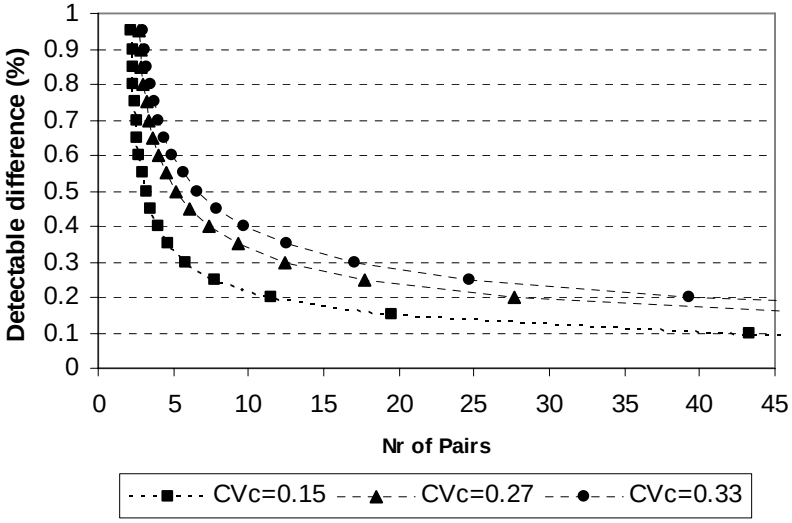
We believe that an improved balance between the arms can be achieved in similar community randomized trials by pair-matching on criteria other than the outcome. Unless a sufficiently long baseline follow-up period is envisaged, the incidence rates may produce different pairs depending on when the data is collected, because diarrhoea is a time fluctuating disease.

Based on our experience we recommend characterizing the randomization units (communities) by baseline potential confounding factors. For instance, proportions of children per age groups < 5 years, socio-economy status, main source for drinking water, hygiene behaviour, use of other disinfection methods, water management and consumption habits. Communities can be then placed in pairs according to similarities in those characteristics. Randomization within balanced pairs will follow reducing thus the risk of bias due to baseline differences. In order to assess similarity in terms of all the baseline characteristics, multivariate exploratory techniques can be applied. One example is given in chapter VI, where similarities between households were explored using 6 indicators of SODIS use, and households were grouped according to their multivariate resemblance. Other approach consists of estimating the probability of receiving the treatment conditioned on similarities between communities in terms of the baseline characteristics. The estimating method is a logistic regression where communities would be paired depending on the similarities in their conditional probabilities or scores. This method is called *Propensity Scores* and is widely used to reduce bias due to confounding in observational studies [3, 4].

The trial was powered to estimate a 33% reduction of the diarrhoea incidence presuming 5 episodes/child/year in the control arm. The simulations of chapter III suggest that all the analytical methods were able to detect an effect of this magnitude considering: 20 clusters, a moderate clustering, cluster size imbalance. However, the estimated effect was lower, i.e. a 19% reduction (RR = 0.81, CI: 0.59 – 1.12). Moreover, the level of clustering was higher than the one assumed during sample size calculations (between-cluster coefficient of variation CV_c = 0.27, CI: 0.11 – 0.46).

These findings provide valuable information for the design of new trials. In this context, we conducted post-hoc power calculations not in order to determine the current "likely state of nature" [5] but to evaluate how well future trials can be conducted given the set of plausible situations we found and the design we plan to implement. Results are displayed in Figure VII.1. The measures of clustering are taken from the main report (chapter II), based on GLMM and Bayesian analyses (chapter IV). But confidence limits are reduced to a more realistic range. Assuming the clustering found in BoliviaWET (CV_c = 0.27), the post-hoc calculations suggest that 28 pairs would be required to detect a 20% reduction in diarrhoea incidence with 80% power. This represents ≅ 2.5 greater sample size to detect a reduction in one episode/child/year from 5 in the control arm.

Figure VII.1: Expected detectable difference with an 80% power at three between-cluster coefficient of variation (CV_c).



The calculations assume the average number of persons-years per cluster observed in BoliviaWET (33 per cluster). Note that adjusting the number of participants per cluster would improve power only if clustering is low. Intuitively, high clustering implies high similarity among individuals within clusters, in which case increasing the number of individuals per cluster would not really help. This fact is well illustrated in Figure 1 from reference [6]. Reliable estimates of the clustering level are thus required.

The CONSORT statement in its extension to CRTs highlights the importance of reporting intra-cluster correlation estimates along with confidence limits [7, 8]. From the two measures of clustering, the intra class correlation coefficient ρ and CV_c, we devote chapter IV to methods for point and interval estimation of CV_c. The choice of the latter is rooted in the fact that CV_c is straightforwardly obtained when modelling count data. Based on asymptotic properties (see equation (9) chapter IV) CV_c is approximately equal to the root of the random-effect variance of a random-intercepts model with log link function ($\sigma_c \cong CV_c$). Another advantage is that overdispersion can be simultaneously modelled by specifying distributions that account for it (e.g. Negative Binomial) when using GLMM methods.

Our findings point out that GLMM with NB distribution or similar Bayesian hierarchical models provide the best point estimates of CV_c. The latter with a conservative (upward) bias when the underlying $CV_c < 10\%$ (Figures III.3 and IV.1), but with the best performance in terms of interval estimation. We also found that for overdispersed counts, the cluster-level point estimating method of CV_c (outlined in [6]) may seriously overestimate clustering when the underlying $CV_c \leq 25\%$ (Figure IV.1).

The estimation of CV_c for the BoliviaWET data, suggested that the Poisson assumption may lead also to overestimating CV_c if the outcome is overdispersed. The magnitude of the bias observed in BoliviaWET data was comparable to that of the cluster level method (Table IV.2). We believe that the unexplained Poisson variability went to σ_c^2 making the estimate grater. Conversely, Poisson models with overdispersion corrections (where the variance function $v(\mu)$ was replaced by $\phi v(\mu)$),

provided low CV_c estimates, comparable to analyses where outliers were excluded (Table IV.2). For this reason, we recommend σ_c from GLMM or Bayes-HM as the estimate of CV_c, with the CI extracted from the Bayesian posterior distribution. NB distribution is recommended to handle overdispersion and seems to be reasonably conservative in the presence of extreme observations.

It has been suggested that CV_c below 0.25 often occurs in real field trials and the value rarely exceeds 0.50 [6]. Our estimate of CV_c = 0.27 (CI: 0.11 – 0.46) from BoliviaWET is an important finding that adds to the knowledge of this indicator. The confidence limits give an idea of the uncertainty and imprecision of CV_c. They can be used in sensitivity analysis of sample size calculations to different CV_c over a plausible range. For example, values of CV_c between the point estimate and a plausible upper limit can be simulated for different sample sizes. The ultimate sample size will reflect the extent to which the investigator wishes to guard against underestimating the required sample size, provided that the upper 95% limit might suggest an infeasible large sample size [9].

A final consideration concerning sample size calculations regards the effect of cluster size imbalance. In chapter III, we found that high imbalance (coefficient of variation of cluster size = 60%) affected the performance of the cluster level t-test and the individual level GEE analysis. In line with our findings, imbalance was also reported elsewhere to influence power and consequently required sample size [10-12]. A cluster size variation > 23% will be enough to affect power in CRT [12]. We therefore recommend accounting for cluster size variation in order to avoid the underestimation of sample size.

Analysis of CRTs.

Consistent with literature on continuous and binary data [13-17], our results show that random-effect (RE) methods are preferable to GEE and cluster level analysis for overdispersed counts under field trials situations. We simulated trials with 10, 20 and 40 clusters in total, different clustering levels (CV_c = 0.05, 0.15 and 0.40) and cluster size imbalance (balance, slightly imbalance and highly imbalance). The methods

compared were: the t-test of cluster-level incidence rates, GEE with empirical and model-based variance estimators, GLMM and Bayes-HM. Below we present some reflections on our overall findings.

The five methods produced accurate RR estimates during the simulations (Figure III.1). The bias was rather small (generally < 3%), but greater (4% – 6%) for all the methods when clustering was high and a sample size was 10 clusters. The stability of the RR was similar across the methods and was primarily influenced by clustering and sample size. The analysis of BoliviaWET, however, provided evidence that the t-test RR may yield biased conservative RR by ignoring the existence of extreme disease responses concentrated in one of the trial arms (TableIII.3, Figure III.5).

As remarked already in chapter I, methods using cluster-level summary statistics may be inefficient since they ignore the within-cluster variation and cluster size [12, 18]. It implies the disregard of imprecision of each summary statistics, which may be considerable in practice. Our experience from the simulation study on overdispersed counts, and the analysis the BoliviaWET data confirmed it. Very high coverage probabilities as a result of wide but unstable CI were found for the t-test during the simulations (Figure III.2, Table III.1). Versions of cluster level methods weighting by cluster size, or within cluster variance are known to improve efficiency [19, 20].

Some approaches have been reported to deal with the impossibility of cluster level methods to adjust for individual covariates. Cluster level t-tests performed on Poisson residuals from a regression that previously adjusts for covariates have been proposed [21]. Some cluster-level methods may be attractive to estimate effects at the risk difference, risk ratio, or odds ratio scales of unadjusted analysis of binary data, because they are easy to calculate [20].

Our results suggest considering GEE for CRT analysis with caution if the trial has less than 40 clusters in total. Narrow CI, anticonservative coverage and high chances of falsely significant results are expected. The reasons are the underestimation of standard errors (SE) by the robust variance estimator, already discussed in chapter III. GEE with model-based variance estimators produced almost identical unfavourable results, suggesting problems with the specification of the working correlation

structure. In addition, high cluster size imbalance reduced GEE coverage, when sample size ≤ 20.

Recent research on GEE provides tools for enhancing the method under the situations studied here. Although not yet implemented in standard statistical software, bias correcting methods are described for amending SE underestimation [22-25]. Simpler modifications regarding the use of the t-distribution rather than z have also been studied and proved to achieve nominal coverage in small samples [19]. Additionally, methods to identify or implement alternative correlation structures have been described elsewhere [26, 27]. We believe that GEE are potentially attractive in CRTs because of their desirable population average interpretation of the intervention effect.

Random effect models via restricted pseudo-likelihood or MCMC yielded stable CIs, nominal coverage and nominal type I error rates (chapter III). This behaviour was robust to sample size, clustering and cluster size imbalance. We warn however that such desirable performance is subject to the fulfilment of the model assumptions [28]. The impact of misspecification of the outcome variance was evident when analysing the BoliviaWET data (Table III.3). Furthermore, the misspecification of the random-effects distribution is known to seriously bias the estimates of the variance of the random-effects (σ_c^2). This has secondary effects on SEs, CIs and the hypothesis testing behaviour of the fixed-effects structure in the model [29, 30].

We recommend therefore RE analysis for community randomized trials with ≤ 40 clusters. For count data, the RR would have both CS and PA interpretations [31]. However the appropriate estimation of clustering would depend upon the number of clusters. While < 6 levels are considered unreliable for variance component estimation [28], we found that even 10 clusters were insufficient to avoid highly unstable estimates (Figures III.3, IV.1). In terms of methods for parameters estimation in RE models, pseudo-likelihood may produce bias in situations with small number of individuals per cluster [15]. Numerical integration and Bayesian analysis via MCMC were shown to have a better performance than pseudo-likelihood in complex design situations [15, 32]. The flexibility of the Bayesian analysis provides other remarkable advantages. Full posterior distributions of the model parameters, and of other

quantities not directly specified (e.g. CV_c, ρ), allows reporting uncertainty measures even for quantities where standard solutions may not exist. The use of prior knowledge can be also seen as a gain, for instance, in the case of intra-cluster correlation.

Finally, RE models are more flexible in analysing complex designs (e.g. nested hierarchies of more than 2 levels, pair-matching, repeated cross-sectional studies). Implementation is undemanding now with the GLIMMIX procedure in SAS, the GLLAMM procedure in STATA, the lme4 library in R or the MLwiN software to mention a few. For models with random-effects that are not normally distributed, implementation is possible via H-likelihood [33], or Bayesian hierarchical models in Winbugs.

Overdispersion.

The Poisson model is almost always considered for analysing count data. It implies equidispersion, i.e., the mean of the response equals its variance. Unfortunately, this assumption is seldom met in practice. Overdispersion, defined as the extent to which the variance exceeds the mean, occurs more often when the responses are correlated, or by an excess of variation between response probabilities or counts [34]. The consequences of ignoring overdispersion in statistical modelling are the underestimation of SE and misleading inference for the regression parameters.

We detected a high level of overdispersion in the BoliviaWET data, even after accounting for intracluster correlation with a Poisson random-effect model (Table I.3). The specification of the NB distribution remarkably improved the fit and handled overdispersion appropriately (Table I.3, Figure I.3). As already pointed out, NB can be viewed as a special form of Poisson, where the mean parameter is a random gamma distributed variable (Poisson-Gamma mixture), whereas the overdispersion correction $\phi v(\mu)$ is merely an inflation of the Poisson variance [34, 35]. We therefore believe that NB models address overdispersion in a more natural manner than just correcting the variance $v(\mu)$ by $\phi v(\mu)$. In the context of CRTs, a NB model viewed as a Poisson-gamma mixture with normally distributed cluster random effects is

equivalent to $Y \sim$ Poisson(μ) where $\mu = \lambda\delta = \exp(x\beta + \xi)$, $\lambda \sim \Gamma(\alpha,\beta)$, $\xi \sim N(0, \sigma_c^2)$. The model parameters can be estimated via full maximum likelihood or MCMC [36]. Alternatively, the NB model could be derived as a GLM with cluster random effects, with parameters estimated via restricted pseudo likelihood [34, 37]. Other models may be also adequate in case of deviations from the equidispersion assumption. Some examples are the Zero-inflated Poisson, Zero-inflated NB in case of excessive zero counts, or Zero-truncated NB when zero counts are structurally excluded from the model [34]. We encourage to regularly verify the residual overdispersion and to fit the model that best fits the data.

Analysis of the outcome and the intervention

Exploratory techniques and statistical modelling were combined to answer specific questions in chapters V and VI. Does the term *k'echalera* employed to report diarrhoea in Quechua speaking settings in rural Bolivia correspond to the standard definition of diarrhoea?. Which are the factors that influenced in the adoption of SODIS in the intervention arm of BoliviaWET?.

Multiple correspondence analysis (MCA) for categorical data or principal component analysis (PCA) for quantitative data, were applied to explore the multivariate patterns of similitude among observation units [38]. MCA on the Burt matrix, contributed to confirm the association of diarrhoeal symptoms among themselves and with the reports of *k'echalera* (Figure V.1). A MCA on the Binary matrix displaying the distribution of child-days of observation showed the similitude of responses given to the questionnaire confirming such associations at individual level.

An in-depth analysis of the relation of the diarrhoeal symptoms and the vernacular term gave lights on the perception of diarrhoea of rural Bolivian mothers in terms of the combination of symptoms that may predict *k'echalera*. Some differences with the standard definition were found and both the sensitivity and specificity of *k'echalera* were estimated using Bayesian modelling assuming imperfect gold standard. We believe that the differences found provide the motivation to evaluate the validity of the standard definition in settings where cultural aspects, nutrition habits and

environment may be responsible of a differentiation between the true changes in defecation patterns and the world diarrhoea definition.

In chapter VI, we investigated the ways of identifying a plausible and objective indicator of SODIS adoption in the intervention arm of BoliviaWET. Four indicators of use and two of monitoring were identified as to quantify SODIS adoption from different perspectives. We wished to differ from the regular ways of quantifying SODIS adoption via self reports or a sole indirect measurement (e.g. presence of SODIS bottles on the roof).

The households were compared in terms of the six indicators via PCA. The similitude among households was visualized in a space conformed by the first 3 principal components (Figure VI.1.b). This exploratory tool allowed us to 1) interpret the patterns of response to the six indicators 2) identify the existence of possible groups of users and 3) to validate the identification of typologies of SODIS-user groups obtained by grouping hierarchical methods based on the 6 indicators [38]. The final typologies resulted in five groups of households, with similar households within groups in terms of all the indicators and different to households from other groups.

We believe this approach is superior to others where the first principal component (PC) is selected as an index that summarises the variation of the variables of interest [39]. Our approach accounts for the information of all the variables simultaneously while the first-PC approach would account only for the subset of variables that describe it. A further disadvantage of the latter is that the ranking of individuals by the scores defined by the first PC is only interpretable for the variables in the subset that have a high linear relation with it.

The application of hierarchical classification methods (cluster analysis) was performed assessing different metrics of similitude and evaluating the several grouping algorithms. Again, the method provided a meaningful classification because we validated the algorithms performance visualizing the grouping results in the PCA data cloud. We warn that a blind application of both PCA and cluster analysis may produce misleading results if the true patterns of variables and individuals relationships are not explored and properly interpreted.

Following the SODIS-users definition, we estimated the effect of household-related and community-level factors on the chance of a household belonging to one of the identified SODIS-user groups. Multinomial regression was applied and within-community correlation of households was allowed for by introducing random effects. The descriptive results show already clear associations between a set of factors with the SODIS-user groups. The multinomial model, while showing similar suggestive tendencies, is unable to detect significance in some cases. We believe that our findings are substantially persuasive for SODIS dissemination programs, although the analysis of 11 clusters in such a complex RE multinomial model is likely underpowered for hypothesis testing in such hierarchical model.

Many topics for statistical research remain open concerning our experience handling and analysing BoliviaWET data. The mail bulk of this thesis deals with CRTs with a completely randomized design. Analytical issues and estimation of clustering measures from pair-matched designs were not addressed. Although we introduced the notion of overdispersion in the analysis, we did not report formally the effects of different magnitudes of overdispersion in our simulations. Another topic that also appealed our attention was the study of imputation methods for diarrhoea for individual days, as a function of diarrhoea occurrence during past days or weeks.

In conclusion, the simulation studies suggest that GLMM and Bayesian models are appropriate for the analysis of overdispersed count data in CRTs in sample sizes ≤ 40 clusters in total. The estimation of the between-cluster coefficient of variation via GLMM and Bayes-HM is also appropriate. The Poisson model may seriously bias both the RR and CV_c estimates. The NB model with normal random-effects provides a natural way to address overdispersion of count data in a CRT. We, encourage to check the residual overdispersion and to apply the (Poisson or extra-Poisson) model that best fits the data.

The BoliviaWET trial found no strong evidence of reduction of the diarrhoea incidence in children < 5 years in families using SODIS. In terms of secondary analyses, we conclude that the vernacular term *k'echalera* does refer to a change in

the regular stool patterns associated with diarrhoea, although it differs from the symptoms-based diarrhoea definition in some aspects. We found that intensity of exposure to the SODIS campaign, latrine ownership, lack of electricity, and having severely wasted children living in the home are associated with uptake of SODIS.

2. References.

1. Klar N, Donner A. The merits of matching in community intervention trials: A cautionary tale. *Statistics in Medicine* 1997; **16**(15):1753-1764.

2. Murray DM. *Design and analysis of group-randomized trials*. Oxford University Press: New York, Oxford, 1998.

3. D'Agostino RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 1998; **17**(19):2265-2281.

4. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine* 2008; **27**(12):2037-2049.

5. Hoenig JM, Heisey DM. The abuse of power: The pervasive fallacy of power calculations for data analysis. *American Statistician* 2001; **55**(1):19-24.

6. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *International Journal of Epidemiology* 1999; **28**(2):319-326.

7. Campbell MK, Elbourne DR, Altman DG. The CONSORT statement for cluster randomised trials. *Medicina Clinica* 2005; **125**:28-31.

8. Campbell MK, Grimshaw JM, Elbourne DR. Intracluster correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. *BMC Med.Res Methodol.* 2004; **4**:9.

9. Ukoumunne OC, Davison AC, Gulliford MC, Chinn S. Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient. *Stat.Med.* 2003; **22**(24):3805-3821.

10. Kerry SM, Bland JM. Unequal cluster sizes for trials in English and Welsh general practice: implications for sample size calculations. *Stat.Med.* 2001; **20**(3):377-390.

11. Guittet L, Ravaud P, Giraudeau B. Planning a cluster randomized trial with unequal cluster sizes: practical issues involving continuous outcomes. *BMC.Med.Res.Methodol.* 2006; **6**:17.

12. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International Journal of Epidemiology* 2006; **35**(5):1292-1300.

13. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: A review of recent methodological developments. *American Journal of Public Health* 2004; **94**(3):423-432.

14. Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials and Statistics in Medicine. *Statistics in Medicine* 2007; **26**(1):2-19.

15. Heo M, Leon AC. Comparison of statistical methods for analysis of clustered binary observations. *Statistics in Medicine* 2005; **24**(6):911-923.

16. Heo M, Leon AC. Performance of a mixed effects logistic regression model for binary outcomes with unequal cluster size. *Journal of Biopharmaceutical Statistics* 2005; **15**(3):513-526.

17. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? *Stat.Med.* 2009; **28**(2):221-239.

18. Omar RZ, Wright EM, Turner RM, Thompson SG. Analysing repeated measurements data: A practical comparison of methods. *Statistics in Medicine* 1999; **18**(13):1587-1603.

19. Ukoumunne OC, Carlin JB, Gulliford MC. A simulation study of odds ratio estimation for binary outcomes from cluster randomized trials. *Statistics in Medicine* 2007; **26**(18):3415-3428.

20. Ukoumunne OC, Forbes AB, Carlin JB, Gulliford MC. Comparison of the risk difference, risk ratio and odds ratio scales for quantifying the unadjusted intervention effect in cluster randomized trials. *Statistics in Medicine* 2008; **27**(25):5143-5155.

21. Bennett S, Parpia T, Hayes R, Cousens S. Methods for the analysis of incidence rates in cluster randomized trials. *International Journal of Epidemiology* 2002; **31**(4):839-846.

22. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001; **57**(1):126-134.

23. Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 2001; **96**(456):1387-1396.

24. Guo X, Pan W, Connett JE, Hannan PJ, French SA. Small-sample performance of the robust score test and its modifications in generalized estimating equations. *Statistics in Medicine* 2005; **24**(22):3479-3495.

25. Lu B, Preisser JS, Qaqish BF, Suchindran C, Bangdiwala S, Wolfson M. A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics* 2007; **63**(3):935-941.

26. Hammill BG, Preisser JS. A SAS/IML software program for GEE and regression diagnostics. *Computational Statistics & Data Analysis* 2006; **51**(2):1197-1212.

27. Hin LY, Wang YG. Working-correlation-structure identification in generalized estimating equations. *Stat.Med.* 2009; **28**(4):642-658.

28. Brown H, Prescott R. *Applied Mixed Models in Medicine*. Wiley & Sons: Ontario, 2001.

29. Litiere S, Alonso A, Molenberghs G. Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics* 2007; **63**(4):1038-1044.

30. Litiere S, Alonso A, Molenberghs G. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine* 2008; **27**(16):3125-3144.

31. Young ML, Preisser JS, Qaqish BF, Wolfson M. Comparison of subject-specific and population averaged models for count data from cluster-unit intervention trials. *Statistical Methods in Medical Research* 2007; **16**(2):167-184.

32. Localio AR, Berlin JA, Ten Have TR. Longitudinal and repeated cross-sectional cluster-randomization designs using mixed effects regression for binary outcomes: Bias and coverage of frequentist and Bayesian methods. *Statistics in Medicine* 2006; **25**(16):2720-2736.

33. Lee Y, Nelder AJ, Pawitan Y. *Generalized Linear Models with Random Effects*. Chapman & Hall: Boca Raton, 2006.

34. Hilbe J.H. *Negative Binomial Regression*. Cambridge University Press, New York: 2007.

35. SAS Institute Inc. *SAS/STAT 9.1 user's guide*. SAS institute Inc.: Cary: NC, 2004.

36. Molenberghs G, Verbeke G, Demetrio CGB. An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis* 2007; **13**(4):513-531.

37. SAS Institute Inc. *The GLIMMIX Procedure*. SAS Institute Inc.: Cary, North Carolina, USA, 2006.

38. Lebart L, Morineau A, Piron M. *Statistique exploratoire multidimensionnelle*. Dunod: Paris, 2000.

39. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data - Or tears: An application to educational enrollments in states of India. *Demography* 2001; **38**(1):115-132.

Appendices

Appendix A. SAS codes for implementing a GLMM analysis on NB count data following both the pair-matched and completely randomized (random-intercepts) designs.

Assuming pair-matching

y : nr of episodes per child.
Intervention : 1, 0 (SODIS, Control)
Dayatrisk : nr of days at risk
Pair : 1,2,…,11
Cluster_pair : 1, 2.

1. GLMM specification.

```
proc glimmix data = dataset;
      lnrisk=log(dayatrisk);        *logarithm of the FU-time;
      class pair cluster_pair;
      model y=intervention/
            dist=negbin       *NB distribution;
            link=log          *log link function;
            offset=lnrisk     *log(FU-time);
            ddf=10      *denominator df for testing H0:β=0 (11-1)*(2-
1);
            cl                *displays the CI;
            solution;         *displays the parameter estimates;
      random pair pair*cluster_pair;      *Specifies the between-
pairs and within-pairs random effects;
run;
```

2. Multilevel regression specification.

```
proc glimmix data = analysis;
      lnrisk=log(dayatrisk);
      class pair cluster_pair;
      model y=intervention/
            dist=negbin
            link=log
            offset=lnrisk
            ddf=10
            cl
            solution;
      random int cluster_pair /sub =pair;
run;
```

Ignoring pair-matching (random-intercepts model)

Y : nr of episodes per child.
Intervention : 1, 0 (SODIS, Control)
Dayatrisk : nr of days at risk
Cluster : 1,2,…,22

1. GLMM specification.

```
proc glimmix data = table3;
      class cluster;
      lnrisk = log(dayatrisk);
      model y = intervention /
                        dist = negbin
                        link = log
                        offset = lnrisk
                        ddf = 20
                        cl
                        Solutions;
      random cluster;
run;
```

2. Multilevel regression specification.

```
proc glimmix data = table3;
      class cluster;
      lnrisk = log(dayatrisk);
      model y = intervention /
                        dist = negbin
                        link = log
                        offset = lnrisk
                        ddf = 20
                        cl
                        Solutions;
      random int /sub = cluster;
run;
```

Appendix B. SODIS Promotion and Implementation Scheme

Time		11 control community-clusters (222 households)		11 intervention community-clusters (262 households)		
		Community level	Household level	District level*	Community level	Household level
6 weeks baseline	(Nov - March 05)	Baseline, pair-matching of control and intervention communities based on the diarrhoea incidence in children <5 years old, random assignment of the SODIS intervention to one of the community pairs§				
3 months before starting one-year follow-up	(April - June 05)			A	A B	
Follow-up: 1-6 months	(June - Dec 05)		G	C	A B D	G E F
Follow-up: 7-12 months	(Jan - June 06)		G	C	A B D	G E F
Intervention: 3 months	(July - Sept 06)	A B D	E			

* Municipality, health - and school system

Symbol	Promotion activity	Content, topic number
A	Introduction to and consolidation of SODIS and related water, sanitation, hygiene, and health issues	1; 2; 3; 4; 5
B	Community event (monthly)	1; 4; 5; 6
C	School event (two monthly)	1; 8
D	Motivational micro-project	7
E	Two weekly household visits	Address day-to-day problems with SODIS-application and management. No specific hygiene messages.
F	Weekly SODIS monitoring	Observational; by community-based staff (independent from the SODIS-implementing NGO)
G	Health monitoring	Health diary kept by mothers; collected by community-based staff (independent from the SODIS-implementing NGO)

Appendix C. Eigenvectors, eigenvalues and correlation coefficients of the first three principal components (Z) of 6 indicators of SODIS-use

Indicator	Z_1		Z_2		Z_3	
	e	**r**	**e**	**R**	**e**	**r**
1. Bottles sun-exposed	**0.57**	**0.87**	-0.15	-0.21	-0.08	-0.08
2. Bottles ready to drink	**0.52**	**0.80**	-0.09	-0.13	-0.30	-0.31
3. Classified user	**0.58**	**0.89**	-0.19	-0.27	0.17	0.18
4. Time behavioral change	0.05	0.08	-0.19	-0.27	**0.92**	**0.94**
5. Time in study (Tool 1)	0.20	0.30	**0.67**	**0.93**	0.12	0.12
6. Time in study (Tool 2)	0.18	0.28	**0.67**	**0.94**	0.13	0.14
Eigenvalue	2.38		1.96		1.05	
Cumul. explained variance (%)	39.7		72.3		89.9	

e = Eigenvector
r = Pearson correlation coefficient

Curriculum Vitae

Name:	**Gonzalo Durán Pacheco**, PhD
Address:	Bäumliweg 18
	4125 Riehen, Switzerland
Tel:	+ 41 61 687 35 94 (office)
Email:	**gonzalo.duranpacheco@gmail.com**
Date of Birth:	3rth of January 1977
Place of Birth:	Potosí, Bolivia
Marital Status:	Married

Education

February 2007 - 2009	PhD in Public Health and Epidemiology Department of Public Health and Epidemiology, Swiss Tropical Institute (STI) University of Basel, Basel, Switzerland **PhD Item: Biostatistics** PhD Title: Analysing cluster randomized trials with count data by Frequentist and Bayesian methods. The BoliviaWET trial: Assessing the effect of SODIS on childhood diarrhoea, (Grade: *Summa Cum Laude*)
September 2001 - October 2002	Diplôme d'Etudes Supérieures (equivalent to Master's degree) Institut de Statistique et Recherche Opérationnelle, Université Libre de Bruxelles, Bruxelles, Belgique **Master Item: Statistics and operations research** Degree project: "Application of some statistical methods in a cross-over preference study", Stay report (Merck Sharp & Dohme, Europe Inc.) (Grade: *Distinction*)
February 1994- December 2000	Licenciatura en Biología (**B.Sc. in Biology**) University of San Simón (UMSS), Cochabamba, Bolivia

Training

September – October 2002	Trainee in **pharmaceutical statistics** Company: Merck Sharp & Dohme, Europe Inc. City: Brussels Country: Belgium

Employment

December 2003 – February 2007	**Trial statistician & coordinator of the data management unit** in a cluster randomized trial on solar water disinfection (SODIS) in rural Bolivia (BoliviaWET, funded by the US-National Institutes of Health (NIH)). Main tasks: coordinating of data management, data quality control and leading the primary/secondary statistical analysis within an international team of scientists (STI/University of California at Berkeley/UMSS)
March 2005 and January - March 2006	**Statistical Consultant:** Two consultancies in statistical analysis (time series data): The effect of atmospheric contaminants on health. Swiss agency for development and cooperation, Project Clean Air. Cochabamba Bolivia

November 2004 – January 2005 and July – November 2005	**Statistical Consultant:** Two medium-term consultancies on study design, coordination of the data management unit, and statistical analysis: Socio-economical impact study of the cooperation projects PRODEVAT and PRAEDAC (European Commission) in the provinces Arque/Tapacari and Chapare, Cochabamba Bolivia PRODEVAT, PRAEDAC projects (EU), Cochabamba Bolivia
January 2003 – October 2004	**Statistical Consultant:** 8 short/medium term consecutive consultancies in study design, data management and statistical analysis of survey data. Development Alternatives Inc., Cochabamba Bolivia
January 1999 – November 2006	**Statistical advisor:** Advisor in sampling/experimental design, data management and statistical analysis of four research centres of the school of Biology, San Simon University.

Scientific Publications and poster presentations

Duran Pacheco G., Hattendorf, J. Colford, Jr. J. Mäusezahl D., & T. A. Smith. Point and Interval estimation of the between-cluster coefficient of variation for overdispersed counts in cluster randomized trials. Working manuscript.

Duran Pacheco G., Christen A, Arnold B, Hattendorf J, Armaza A, Colford, Jr., T. A. Smith, J. Mäusezahl D. Reporting diarrhoea through a vernacular term in Quechua speaking settings of rural Bolivia. Working manuscript.

Christen A, **Duran Pacheco G**., Hattendorf, J., Cevallos M., Morante C., Arnold B., Colford Jr. J., Mäusezahl D. Implementing Solar Water Disinfection technology: Factors influencing household adoption of SODIS in a community randomized trial in Bolivia. Working manuscript.

Duran Pacheco G., Hattendorf, J. Colford, Jr. J. Mäusezahl D., & T. A. Smith. Performance of Analytical methods for overdispersed counts in cluster randomized trials: sample size, degree of clustering and imbalance. Statistics in Medicine 2009; 28:2989-3011.

Mäusezahl D., Christen A., **Duran Pacheco G**., Alvarez Tellez. F., Iriarte M., Zapata M.E., Cevallos M., Hattendorf J., Daigl C. M., Arnold B., T. A. Smith & J. M. Colford, Jr. A cluster-randomized, controlled trial of solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia. PlosMed 2009, DOI:10.1371/journal.pmed.1000125.

Terrazas A. F. Baudoin, J. P. & **G. Duran Pacheco** (2007) Procesos dinámicos locales para la conservación in situ de la diversidad genética de tubérculos andinos cultivados en el microcentro de Candelaria (Cochabamba, Bolivia). *Plant genetic Resources Newsletter, No. 152: 1-11*

Mäusezahl D., Christen, A., Niggli M., Hobbins M., Daigl M., **Duran Pacheco G**, Romero A. M., Iriarte, M. Estrella M. & J. Colford. 2006 Pure drinking water from sunlight: Solar disinfection water can prevent gastrointestinal diseases in rural Bolivia. Poster in the 12th International congress on Infection diseases.

Duran Pacheco G. (2006) Discrete Probability Models to Assess Spatial Distribution Patterns in Natural Populations and an Algorithm for Likelihood Ratio Goodness of Fit Test. *Acta Nova, vol. 3 nr 3: 543-563*

Duran Pacheco G; X. Cadima & J. Zeballos (2004) Desarrollo de una Colección Núcleo de la Colección de Papa Cultivada (Solanum Ssp.) del Banco de Germoplasma de Raíces y Tubérculos Andinos de Bolivia. *In XI International congress of Andean crops, Cochabamba Bolivia: A -13.*

Duran Pacheco G; N. Sotomayor; T. Ávila Alba & C. Rocabado (2004) Efecto de un Complejo Vitamínico y de la Densidad de Plántulas en el Desarrollo in vitro y en Invernadero de la Variedad de Papa Alpha

(Solanum tuberosum, ssp. tuberosum). *In XI International congress of Andean crops, Cochabamba Bolivia: PA – P- 1*

Teaching experience

Lecturer of Biostatistics, Applied statistics and Mathematical Statistics in 7 graduate programs at the UMSS from 2004 – to present.
Lecturer of Experimental Designs, school of Biology, UMSS (winter 2004 - Spring 2006).

Scientific association membership

- Member of the Bolivian Statistical Society, Bolivia

Other Knowledge

Languages	Fluent: English, French and Spanish. Some knowledge: Quechua
Genetic resources	Experience in: • Sampling design and design of experiments. • Data management and evaluation. • Statistical analysis.
Statistical Software	Highly proficient in: • **SAS** (SAS data step, SAS/SQL, SAS/Stat, SAS/Macro SAS/Graph). Experience in • **R, SAS/JMP, Stata, WinBugs (among others)**
Other Software	Latex, MATLAB, Microsoft Word/Excel/ Power point, as well as some other specific software programs

Reviewer in Scientific Journals

Journal of Tropical Medicine an International Health
Transactions of the Royal Society of Tropical Medicine and Hygiene

Leisure Interests

Piano playing
Travelling

References

Marcel Tanner, Professor
Director of the Swiss Tropical Institute

Thomas Smith, Professor
Head of the Biostatistics unit,

Swiss Tropical Institute Address: Socinstrasse 57, CH-4002 Contact:marcel.tanner@unibas.ch Tel: +41 61 284 82 87 Country: Switzerland	Swiss Tropical Institute Address: Socinstrasse 57, CH-4002 Contact: thomas-a.smith@unibas.ch Tel:+41 61 284 82 73 Country: Switzerland
Daniel Mäusezahl, Ph.D, MPH. Senior Scientist, principal investigator of the BoliviaWET study, Swiss Tropical Institute Address: Socinstrasse 57, CH-4002 Contact:daniel.maeusezahl@unibas.ch Tel:+41 79 3450118 Country: Switzerland	**Marc Hallin, Professor** Director of the research unit in mathematical statistics, Institute of Statistics and Operations Research Free University of Brussels (ULB) Campus de la Plaine C.P. 210 Address: B-1050 Brussels Contact: mhallin@ulb.ac.be Tel:+32 2 650 5899 Country: Belgium

List of lectures followed during the PhD

Course number	Course name	University	Duration	Credit Points
12385-01	(1) Biostatistics II (Bayesian statistics)	Univ. Basel	28/03/07 - 4/07/07	4
2250.07	(2) Advanced Methods in Epidemiology: analysis of clustered data and multilevel modeling	Univ. Bern	09/05/07 - 11/05/07	1.5
2007-ss-en-07	(3) English: Speaking in professional and academic context	Univ. Basel	29/03/07 - 12/07/07	3
2008-ss-en-01	(4) English: Academic writing	Univ. of Basel	01/03/08 - 24/05/08	3
19366	(5) STI research seminar	Univ. of Basel	18/02/08 - 25/05/08	1
-	(6) Applied Bayesian statistics in medical research and health care	Univ. of Bern	12/03/08 - 14/03/08	1
-	(7) Practical Bayesian models for the Health Sciences	Univ. of Bern	20/03/08	0.5
	TOTAL			14