



Swiss Federal Institute of Technology Zurich

Seminar for
Statistics

Department of Mathematics

Master Thesis

Spring 2015

Nicolas Bennett

The title of my thesis
which should be split on
several lines if it is too long

Submission Date: August 25th 2015

Co-Adviser: Anna Drewek
Adviser: Prof. Dr. Peter Bühlmann

[B]ecause we tend to reward others when they do well and punish them when they do badly, and because there is regression to the mean, it is part of the human condition that we are statistically punished for rewarding others and rewarded for punishing them — [Kahneman \(2012\)](#).

Preface

First words and acknowledgements. And we add a lot of text to make sure that it spans more than one line, as otherwise it may not show up.

Abstract

This example thesis briefly shows the main features of our thesis style, and how to use it for your purposes.

Contents

Preface	iii
Abstract	v
Contents	vii
List of Figures	ix
List of Tables	xi
List of Abbreviations and Notation	xiii
1 Introduction	1
2 Biological Background	7
2.1 Microbial Host-Cell Infection	7
2.1.1 Viral Infection Mechanisms	8
2.1.2 Bacterial Entry Mechanisms	10
2.1.3 Actin-based Intracellular Motility	12
2.2 Select Microbial Pathogens	12
2.2.1 Bartonella Henselae	12
2.2.2 Brucella Abortus	14
2.2.3 Listeria Monocytogenes	16
2.2.4 Salmonella Enterica	18
2.2.5 Shigella Flexneri	18
2.2.6 Adenovirus	18
2.2.7 Rhinoviruses	18
2.2.8 Vaccinia Virus	19
2.3 RNA Interference	19
3 Mathematical Background	21

CONTENTS

4 Writing scientific texts in English	23
4.1 Basic writing rules	23
4.2 Being nice to the reader	23
4.3 A few important grammar rules	24
4.4 Things you (usually) don't say in English	27
5 Typography	29
5.1 Punctuation	29
5.2 Spacing	30
5.3 Choice of 'fonts'	31
5.4 Displayed equations	31
5.5 Floats	32
6 First Chapter SfS Template	35
6.1 To include a picture	35
6.2 To make a proof	35
6.3 To include R code	35
6.4 Other information	36
A Complementary information	37
A.1 Including R code with verbatim	37
A.2 Including R code with the <i>listings</i> package	38
A.3 Using Sweave to include R code (and more) in your report	38
B Yet another appendix....	39
B.1 Description	39
B.2 Tables	39
Bibliography	41
Epilogue	43

List of Figures

1.1	Relative frequencies of death causes in 2012 by World Bank income groups	3
1.2	Relative frequencies deadly infectious diseases for 2012 by World Bank income groups	3
2.1	A generalized view of the steps necessary for viral–cellular membrane fusion	9
2.2	Zipper and trigger mechanisms for bacterial host-cell entry	11
6.1	Geyser data: binned histogram, Silverman’s and another kernel .	35
6.2	Geyser data: binned histogram, Silverman’s and another kernel .	36

List of Tables

2.1	The Baltimore classification scheme for viruses	10
4.1	Things you (usually) don't say	27
B.1	Test results	39

Abbreviations and Notation

AIDS acquired immune deficiency syndrome.

AMR Antimicrobial resistance.

Arp2/3 actin-related-protein 2 and 3.

Cdc42 cell division control protein 42 homolog.

COPII coat protein complex II, involved in anterograde endoplasmic reticulum (ER)–Golgi transport.

CSD Cat Scratch Disease.

ER endoplasmic reticulum.

ERAD endoplasmic-reticulum-associated protein degradation.

ERGIC ER-Golgi intermediate compartment.

HDT Host directed therapeutics.

HIV human immunodeficiency virus.

ICAM-1 intercellular adhesion molecule 1.

Ipa *Shigella* invasion plasmid antigen.

LPS lipopolysaccharide.

Rac1 Ras-related C3 botulinum toxin substrate 1.

SARS severe acute respiratory syndrome.

Sip *Salmonella* invasion protein.

SopE *Salmonella* outer protein E.

SptP secreted effector protein.

T3SS type III secretion system.

LIST OF ABBREVIATIONS AND NOTATION

- T4SS** type IV secretion system.
- TIR** toll-interleukin-1 receptor.
- Tir** translocated intimin receptor.
- TLR** toll-like receptor.
- VEGF** vascular endothelial growth factor.
- WHO** World Health Organization.

Chapter 1

Introduction

Infectious diseases have played an undeniably important role in human history. With human populations becoming sufficiently aggregated to sustain direct life cycle viral and bacterial infectants around 2000 BC, devastating invasions of a growing number of pathogens started to occur ([Dobson and Carper, 1996](#)).

One of the earliest well documented incidence of a large-scale epidemic is known as the Plague of Athens. Starting in 430 BC and lasting roughly three years, a highly infectious disease killed 75'000 to 100'000 people or 25% of Athen's population. This catastrophic event is attributed either to smallpox, a viral infection with *Variola major* or typhus, caused by *Rickettsia* bacteria ([Littman, 2009](#)).

The bacterium *Yersinia pestis* caused three major plague pandemics in the early and late middle ages, as well as in the late 19th century. Originating in northern Africa in 523 AD and spreading around the Mediterranean basin throughout the years 541–546, the Plague of Justinian is assumed to have killed up to half of the population of affected areas. The effect on cities was disproportionately severe. In Constantinople, for example, an estimated 230'000 people out of 375'000 lost their lives to the disease ([Treadgold, 1997](#)). Returning in the years 1347–1351, known today as the Black Death, a plague pandemic again wiped out around half of Europe's population. Death toll estimates range from 15 to 23.5 million ([Zietz and Dunkelberg, 2004](#)). Leaving behind a grim cultural heritage, this catastrophe had a lasting effect on economic and social structures in Europe. The third large-scale outbreak started around 1855 in southern China and quickly spread to Japan, Taiwan and India again wreaking havoc on the affected population.

Bringing diseases such as smallpox, measles (an infection with the *Measles virus*) and typhus to the Americas during the European invasion of the New World had grave repercussions for the indigenous population, carrying no

1. INTRODUCTION

natural resistance towards the newly introduced pathogens. It is estimated that the population of present day Mexico fell from 20 million to 1.6 million over the course of the 16th century due to multiple disease epidemics, critically contributing to the successful colonization of the new continents ([Dobson and Carper, 1996](#)).

Cholera and influenza are further contagious diseases with high mortality rates, responsible for global epidemics. *Vibrio cholerae*, a bacterium which causes infections of the intestine, became widespread in the early 19th century and caused seven pandemics since, the last of which only started in 1961. Antibacterial treatment of sewage and purification of drinking water greatly help to prevent and contain spreading of the disease but in areas with inadequate sanitation, such as Haiti after the 2010 earthquake, it remains a pathogen difficult to control. The influenza virus causes seasonal epidemics characterized by low lethality rates among people with intact immune systems¹. Irregularly occurring influenza pandemics, initiated by zoonosis of new virus strains, against which no natural immunity exists, however, are accompanied by much higher lethality rates. The most significant such event is known today as the Spanish flu pandemic of 1918, costing the lives of 50–100 million, nearly half of which were young, healthy adults ([Taubenberger and Morens, 2006](#)).

In addition to diseases plaguing humanity for centuries, new ones continually emerge. Human immunodeficiency virus (HIV) is believed to have transferred from non-human primates in the early 20th century and the recent outbreaks of severe acute respiratory syndrome (SARS) and swine flu serve as vivid illustrations of such occurrences.

Despite development of means to treat and prevent many previously devastating diseases, infectious pathogens remain a serious threat to global health. In 2012, an estimated total of 58.3 million people died (20.1% in high, 29.4% in upper-middle, 36.5% in lower-middle and 14% in low income countries). Figure 1.1 partitions the total death count into World Bank income groups and causes. In low income countries, infective diseases are the most prevalent cause of death (39.6%), followed by maternal and perinatal complications with substantial margin (20.8%). In lower middle income countries, cardiovascular conditions catch up (26.5%), but are still almost matched in frequency by infectious diseases (23.3%). In upper middle (8.5%) and high income countries (6.7%), the importance of infectious disease while weakened remains accountable for a significant number of deaths. Globally, infectious diseases are the second most frequent cause of death (18.3%), even

¹In spite of low lethality, these seasonal epidemics still incur significant economic damages. The [World Health Organization \(2003\)](#) estimates annual health care costs and loss of productivity due to influenza at US \$71–176 billion for the United States of America alone.

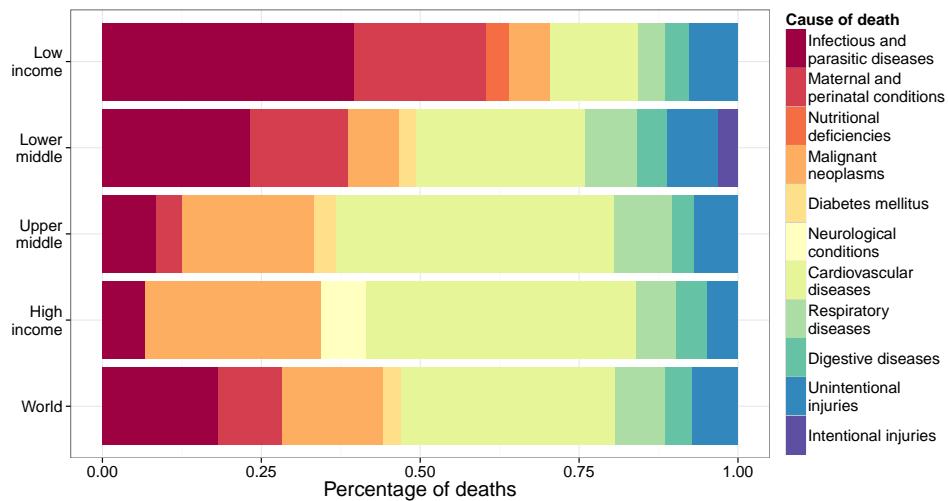


Figure 1.1: Relative frequencies of death causes in 2012 by World Bank income groups. Binning is based on Gross National Income (GNI) per capita and the thresholds are \$1045 or less for low income, \$1046 to \$4125 for lower-middle, \$4126 to \$12745 for upper-middle and \$12746 or more for high income economies. The data was obtained from the [World Health Organization \(2012\)](#).

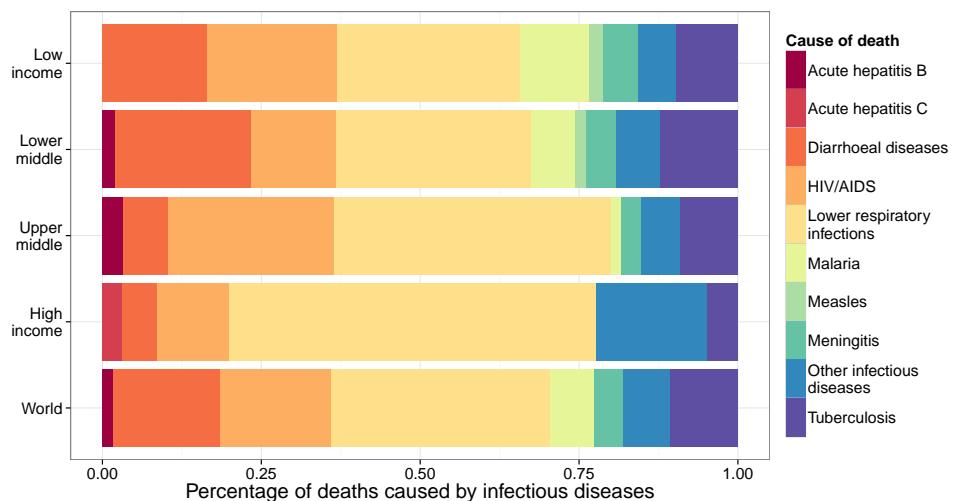


Figure 1.2: Relative frequencies of deadly infectious diseases for 2012 by World Bank income groups. Binning is based on Gross National Income (GNI; see figure 1.1). The data was obtained from the [World Health Organization \(2012\)](#).

more prevalent than all forms of cancer combined (15.8%) and only preceded by cardiovascular diseases (33.7%).

Focusing only on deaths caused by infectious disease, lower respiratory infections are most frequent (for each income region individually, low to high: 28.7%, 30.8%, 43.5% and 57.7% as well as worldwide: 34.5%; cf. figure 1.2). Diarrhoeal diseases and HIV/AIDS are the next most common worldwide

1. INTRODUCTION

(16.9% and 17.3%, respectively) where diarrhea is more prevalent in lower income regions (16.6% and 21.4% versus 7% and 5.6%), while HIV/AIDS plays a major role irrespective of income region (low to high: 20.4%, 13.3%, 26.2% and 11.3%).

Dealing with highly virulent pathogens and preventing their spreading requires a multi-pronged approach. First and foremost, etiology and routes of transmission have to be understood. Knowledge of vectors and natural reservoirs is of great importance as a first line of defense. In the case of plague, for example, insecticides killing fleas were successfully used as a prophylactic measure, as well as controlling rat populations. Sanitary precautions including purification of drinking water, cooking foods well and the usage of disinfectants prevent initial infection, while measures such as sewage treatment, hand washing and wearing face masks help limiting spread among humans. Vaccination is the most important preventive measure. Exposing the immune system to a foreign antigen in a controlled manner artificially induces immunity. Among the great successes of widespread vaccination efforts is the global eradication of smallpox through a coordinated initiative lead by the World Health Organization in the 1970's.

Post-infection therapies include symptomatic treatments, as well as anti-infective drugs. Anti-bacterial or anti-viral agents exploit differences in proteomes between host and pathogen to selectively disable the invader with minimal toxicity to the host. This approach has been tremendously successful throughout most of the 20th century, leading to widespread application and prompting development of resistance towards the commonly used compounds. Adding to the severity of the problem is a lack of discovery of new drugs. No new class of anti-bacterial agents has been found since 1987, causing big pharmaceutical companies to withdraw from the area. The remaining research is mainly focused on improving on existing drugs, leading to a weak product pipeline, especially for the treatment of gram-negative bacteria ([Silver, 2011](#)). In its first global study on anti-microbial resistance, the [World Health Organization \(2014\)](#) notes:

Antimicrobial resistance (AMR) within a wide range of infectious agents is a growing public health threat of broad concern . . . A post-antibiotic era—in which common infections and minor injuries can kill—far from being an apocalyptic fantasy, is instead a very real possibility for the 21st century.

An alternative to pathogen directed search lies in targeting the set of host proteins necessary for infection. Many intracellular parasites subvert cellular functions to gain entry via host-mediated processes such as endocytosis. Upon entry, they move to a suitable niche and rely on host resources for proliferation. Challenges include evading host-cell defense mechanisms, generating sufficient space for replication, nutrient acquisition and keeping the

host alive as long as possible, most of which require complex interactions between invader and host-based mechanisms. Finally, exiting the host cell again requires the parasite to successfully insert itself into existing signaling pathways ([Leirião et al., 2004](#)).

Host directed therapeutics (HDT) offer an escape from the conundrum of wanting to kill but not wanting to select for the surviving microbial parasites. The major challenge under this regimen is finding infectome components that are nonessential for cell survival, as orthogonality of host and infectant can no longer be exploited. Proving feasibility of the approach, [Czyż et al.](#) screened a library of 640 compounds already approved by the United States Food and Drug Administration (USFDA or FDA) for inducing resistance to four intracellular pathogens (*Coxiella burnetii*, *Legionella pneumophila*, *Brucella abortus*, and *Rickettsia conorii*). They found multiple drugs, not classified as antibiotics, that successfully inhibited intracellular bacterial growth while entailing only limited toxicity to host cells (THP-1). [Prussia et al.](#) review the usage of genome-wide screens to study host-pathogen interactions (for HIV and influenza) which in turn serve as basis for rational identification of drug targets for novel host-directed antivirals.

A detailed understanding of the human infectome is of crucial importance to the development of HDT and may even benefit the development of new anti-microbial agents. Feasibility of systematic loss of function screens using RNA interference methodology offers a unique opportunity to investigate complex cellular networks, making this an ideal tool for laying groundwork in combating infectious diseases. Of great importance, however, is ensuring reproducibility and comparability of such datasets, as well as ready availability to the scientific community.

Chapter 2

Biological Background

In order to better understand infectious diseases from a cell biological standpoint, this chapter reviews the current state of knowledge surrounding both bacterial and viral entry mechanisms. A sweeping overview of epidemiology and pathogenesis for several specific bacterial (*Bartonella henselae*, *Bacillus abortus*, *Listeria monocytogenes*, *Salmonella enterica* and *Shigella flexneri*), as well as viral parasites (*Adenovirus*, *Rhinovirus* and *Vaccinia virus*) is given and the chapter concludes with a look at RNA interference as this mechanism is a cornerstone of genome-wide knockdown experiments.

2.1 Microbial Host-Cell Infection

Multi-layered keratinized skin is impenetrable for almost all microbial parasites. Instead they either require breaches such as cuts, scratches, puncture wounds and arthropod bites, or environmental interfaces which offer less impervious protection. Examples include respiratory, gastrointestinal and urogenital tracts, which all contain segments where only a single layer of epithelial cells has to be overcome. Although often protected by chemical defense mechanisms (acidity of the stomach and urogenital tract, as well as microbicidal factors in mucous secretions in the respiratory tract and small intestine), combined with frequent flushing (urination, peristalsis and the coordinated beating of cilia), some microbes have adapted to survive these environments.

For extracellular pathogens to successfully colonize epithelial linings, they must avoid being removed by cleansing mechanisms of the host. Many bacteria accomplish this by expressing adhesins, protein complexes that recognize and bind to specific host-cell receptors, providing host and tissue tropism. Bacterial pili serve to extend reach and penetrate mucous secretions and therefore often carry adhesins. Enteropathogenic *Escherichia coli* have extended this scheme by injecting their own receptor protein Tir (translocated

2. BIOLOGICAL BACKGROUND

intimin receptor) through the T3SS (type III secretion system) into the host cell to which it then attaches. This has the additional advantage that the intracellular domain of Tir can be used to modify host cell behavior ([Alberts et al., 2008](#)).

The outside of many epithelial barriers is covered in natural bacterial flora and crossing over into sterile cavities has the advantage of not having to compete with organisms well accustomed to that particular niche. Furthermore, intracellular pathogens are no longer accessible to antibodies and phagocytic cells and have a nutrient rich environment at their disposals. Mechanisms for entering host cells are described in the following sections.

2.1.1 Viral Infection Mechanisms

Viruses are entirely dependent on host-cell metabolism and machinery for their replication, making them obligately intracellular. The first step of any entering sequence is binding to the target surface. This can be mediated by attachment factors which simply serve to concentrate the virions on the cell membrane or by virus receptors, which additionally act as communicators between host and pathogen. Common attachment factors include glycosaminoglycan chains and sialic acids and are comparatively unspecific. Glycoprotein spikes on enveloped and capsid proteins of non-enveloped viruses provide host specificity by binding cellular receptors. Mostly these cellular receptors serve other purposes which are exploited for infection. Binding affinity for individual interactions may be weak but aggregation of multiple interactions provide virtually irreversible avidity ([Smith, 2012](#)).

Viral import. For viral cell entry, different strategies exist. Enveloped viruses can either directly fuse with the plasma membrane (e.g. HIV) or be endocytosed by the host cell (e.g. influenza), while non-enveloped viruses either create a pore and directly inject their genome into the cytosol (e.g. poliovirus) or are endocytosed (e.g. adenovirus). Endocytosis has major advantages over alternative strategies. Reaching its replicatory niche within the host cell is a difficult task for a microorganism having no means of locomotion and hijacking the endocytic system solves this problem elegantly. Furthermore, maturation of endosomes provides precise environmental cues to the invader for triggering uncoating and release. Both fusion with the cell membrane and injection of viral material into the cytosol leaves back traces of infection to be detected by the immune system. Being completely engulfed by the host, virions leave back no telltale traces. Finally, lytic penetration techniques are not as problematic to the host if only applied inside an endosome.

Many endocytic viruses trigger clathrin-mediated, lipid-raft mediated or macropinocytic mechanisms. The clathrin pathway is most commonly used (for

2.1. Microbial Host-Cell Infection

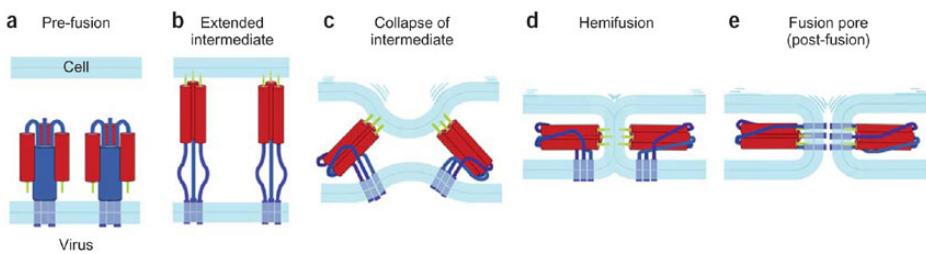


Figure 2.1: A generalized view of the steps necessary for viral–cellular membrane fusion. (a) In pre-fusion conformation, fusion proteins have their hydrophobic fusion moieties tucked away. (b) Reacting to a trigger, such as low pH, the extended intermediate state is formed, characterized by exposed fusion sections, ready to interact with the target membrane. Conformational change in fusion proteins forces the two bilayers into close proximity, yielding first a collapsed intermediate state (c) and subsequently a state of hemifusion (d). Finally a fusion pore is formed (e), stabilized by the post-fusion conformation which is lower in energy than the pre-fusion state. ([Harrison, 2008](#))

example by rhinoviruses, some adenoviruses and influenza), while larger virions such as adenovirus 3 or vaccinia virus initiate macropinocytosis, an actin dependent formation of a lamellipodium which folds back onto the plasma membrane, creating a macropinosome. Lipid raft dependent pathways are poorly understood. Simian virus 40 and polyomaviruses initiate cholesterol dependent formation of lipid rafts and trigger endocytic uptake by inducing plasma membrane curvature.

Upon endocytic uptake, viral pathogens need to uncoat and eject their genetic material into the cytosol, as soon as their replicatory niche is reached. Escape timing is a critical issue, as late endosomes finally turn into lysosomes, capable of digesting their contents. Many enveloped viruses employ fusion mechanisms, which can be classified as type I or type II. For both types, increasing acidity associated with endosome maturation, initiates membrane fusion. Type I fusion proteins are forced into a metastable conformation prior to being added to the viral envelope and low pH triggers a conformational change to a state of lower energy. The energy released is used to force the two membranes close together resulting in their fusion (see figure 2.1). In type II fusion proteins, the critical transformation is not a conformational change but one in quaternary structure.

Non enveloped viruses cannot fuse with host membranes and have developed alternative approaches such as lysis (e.g. adenovirus) or ejecting their genome through pore-forming complexes (e.g. reovirus). Polyomaviruses need to pass through the ER because they rely on ER localized proteins to uncoat their capsid. For export from the ER into the cytosol, they exploit the endoplasmic-reticulum-associated protein degradation (ERAD) pathway, which serves as export mechanism for misfolded proteins from the endoplasmic reticulum to be degraded by proteasomes.

2. BIOLOGICAL BACKGROUND

Table 2.1: The Baltimore classification scheme is based on diversity of genetic system that have evolved in viruses. For each group, a selection of virus families capable of infecting humans, is provided, along with whether the virions are enveloped and the location of their replicatory niche. The data was obtained from [Hulo et al. \(2011\)](#)

	Genome based class	Examples	Enveloped	Replication site
DNA viruses	Group I: dsDNA	Adenoviridae	no	nucleus
		Poxviridae	yes	cytoplasm
	Group II: ssDNA(+)	Parvovirinae	no	nucleus
		Anelloviridae	no	nucleus
RNA viruses	Group III: dsRNA	Reoviridae	no	cytoplasm
	Group IV: ssRNA(+)	Coronaviridae	yes	cytoplasm
		Picornaviridae	no	cytoplasm
		Hepeviridae	no	cytoplasm
	Group V: ssRNA(-)	Filoviridae	yes	cytoplasm
		Paramyxoviridae	yes	cytoplasm
Retro	Group VI: ssRNA(+)-RT	Orthoretrovirinae	yes	nucleus
	Group VII: dsDNA-RT	Hepadnaviridae	yes	nucleus

Replication. Requirements for replication of the viral genome, possible transcription and translation vary widely among different groups of viruses. A classification system, based on the diversity of mRNA production mechanisms has been devised by [Baltimore \(1971\)](#). Table 2.1.1 provides examples of human viruses for each of the 7 groups, whether those viruses are enveloped and where they replicate.

Viral export.

2.1.2 Bacterial Entry Mechanisms

Due to the much larger size of bacterial pathogens, endocytosis is not a feasible mechanism for entry. Phagocytosis, however can deal with particle uptake of this size magnitude. While phagocytosis is a function usually only available to macrophages, some bacteria have evolved mechanisms to induce phagocytosis in other cell types. Furthermore, species such as *Mycobacterium tuberculosis* and *Legionella pneumophila* targeting macrophages have to be able to escape phagosomes or deal with resisting digestion.

Two recurring patterns for inducing phagocytosis in non-phagocytic cells have been described: the zipper mechanism, found in *Yersinia pseudotuberculosis* or *Listeria monocytogenes* and the trigger mechanism used by *Salmonella enterica* and *Shigella flexneri*. Not all entry strategies can be assigned to these two classes and several additional, unrelated pathways have been described.

2.1. Microbial Host-Cell Infection

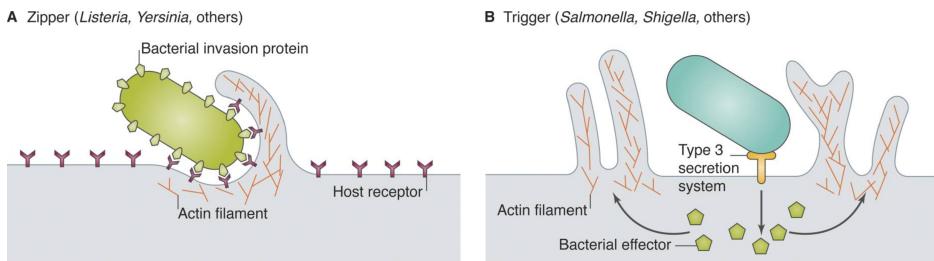


Figure 2.2: Both the zipper (A) and trigger mechanisms (B) are actin dependent and lead to phagocytosis by usually non-phagocytic host cells. Zippering bacteria display an invasion protein on their surface that recruits actin filaments via a host receptor, while triggering bacteria inject an effector into the host cytosol via a type III secretion system leading to uptake. ([Haglund and Welch, 2011](#))

Zipper Mechanism. The first step for zippering bacteria is binding to target cell receptors by expressing adhesins. *Y. pseudotuberculosis* displays invasins on its cell surface, capable of interacting with β_1 integrins, while *L. monocytogenes* uses internalin A, a protein that binds E-cadherin. In both settings, a downstream signaling cascade leads to actin polymerization via recruitment of an Arp2/3 (actin-related-protein 2 and 3) complex and Rac1 (Ras-related C3 botulinum toxin substrate 1), yielding a phagocytic cup.

Cadherin and integrins are usually involved in anchoring cell junctions and the host cell is fooled into thinking a neighboring cell is initiating formation of such a junction. Responding by recruiting actin at the site of bacterial attachment to cover the surface of mistaken invasion proteins leads to engulfment and phagocytosis to the bacterium. This process incurs only modest cytoskeletal rearrangements.

Trigger Mechanism. In case of the trigger mechanism, bacterial T3SS weakly adhere to target cell receptors (CD44 for *Shigella*) and effector molecules are injected into the host cytosol through a pore formed by Sip (*Salmonella* invasion protein) or Ipa (*Shigella* invasion plasmid antigen) proteins. These induce major actin rearrangements that result in localized ruffling of the plasma membrane and subsequent swallowing of the bacterium.

Early in *Shigella* entry, VirA, secreted through the T3SS causes local destabilization of microtubules by binding to α/β -tubulin heterodimers. This in turn stimulates Rac1 (Ras-related C3 botulinum toxin substrate 1) activity, Cdc42 (cell division control protein 42 homolog) recruitment and subsequent Arp2/3 activation leading to protrusions formed by actin filaments. IpaC recruits the Src tyrosine kinase further enhancing actin dynamics. Upon closing of the phagocytic pocket, the T3SS-secreted protein IpaA binds to vinculin and induces actin depolymerization.

Salmonella inject the proteins SopE (*Salmonella* outer protein E) and SptP (se-

2. BIOLOGICAL BACKGROUND

creted effector protein), an activator/inhibitor pair for the GTPase complex Rac1/Cdc42, into the host cell. First, GEF activity of SopE induces massive actin-rearrangements leading to membrane ruffling and facilitating phagocytosis, followed by GAP activity of SptP, restoring the inactive GDP state of Rac1/Cdc42 and leading to actin depolymerization. SopE is degraded more rapidly than SptP, enabling the invading pathogen to reversibly control the pathways exploited for entering.

Other Entry Pathways. In addition to trigger and zipper type uptake, other atypical mechanisms exist. Host cell entry of *Brucella abortus*, for example, has been described as invasome mediated ([Dehio, 2005](#)). In this actin dependent process, bacteria aggregate on the cell surface and trigger their engulfment by injecting bacterial effectors into the host via type IV secretion system (T4SS). The internalized structure is called an invasome. So far, actin has always been the driving force behind internalization. Actin-independent uptake albeit rare, is possible as evidenced by *Campylobacter jejuni*, which have evolved a microtubule dependent invasion strategy ([Kopecko et al., 2001](#)).

2.1.3 Actin-based Intracellular Motility

2.2 Select Microbial Pathogens

A total of 8 viral and bacterial pathogens were studied within the InfectX RTD project by SystemsX. This section shortly describes each organism in terms of microbiological features, pathogenesis, epidemiology and diseases caused in humans. The sections on bacteria are mostly based on [Rolain and Raoult \(2006\)](#) and for viruses, [Craighead \(2000\)](#) was used as a basis.

2.2.1 *Bartonella Henselae*

Bartonella henselae is a short, rod shaped, unflagellated proteobacterium, phylogenetically closely related to the genus *Brucella*, presenting 94.4% 16S rRNA gene sequence homology, compared with *Brucella abortus*. The Gram-negative bacillus is a facultative anaerobic, intracellular parasite and was first described in 1992. Relatively harmless for healthy humans, infections can become life threatening in immunocompromised patients, making the species an important opportunistic pathogen ([Anderson and Neuman, 1997](#)).

Diseases. In immunocompetent humans, infection with *B. henselae* can lead to a condition known as Cat Scratch Disease (CSD). As the name suggests, most patient report being in contact with a cat and transmission often occurs through scratches and bites. Affecting primarily children and young adults (80% are 21 or younger), the self limiting infection typically presents

2.2. Select Microbial Pathogens

itself with lymphadenopathy. Most patients remain afebrile and do not report feeling ill, with low-grade fever and malaise shown in roughly 30% of the cases. Recovery from uncomplicated CSD usually takes 2 to 6 months and requires no specific treatment.

Possible complications include Parinaud's oculoglandular syndrome (granulomatous conjunctivitis in one eye and parotid lymphadenitis on the same side), splenomegaly and hepatic or splenic abscesses, accompanied by fever, weight loss, fatigue and malaise. In 1 to 7% of the cases, the disease spreads to the central nervous system, leading to encephalopathy, but recovery is usually rapid (within several weeks).

Infections with *B. henselae* tend to have more severe consequences for immunocompromised patients, such as bacillary angiomatosis, bacteremia and endocarditis. Acquired immune deficiency syndrome (AIDS) patients suffering from CSD usually experience severe, progressive disease with infection spreading systematically and without appropriate treatment, fatal outcome. *Bartonella spp.* are the only prokaryotes known to be able to induce angiogenic tumors such as bacillary angiomas, which may involve skin, respiratory or gastrointestinal epithelia, heart, liver, spleen, bone marrow, muscles, or lymph nodes. Bacteremia may lead to inflamed heart valves, usually requiring endocarditic patients to have heart valve replacement surgery.

Pathogenesis. *B. henselae* are capable of intracellular growth in both epithelial cells, and erythrocytes. Bundle forming type IV pili are essential for binding to target cells, making them important virulence factors. Internalization into red blood cells may be spectrin mediated and the bacterial protein deformin may also be involved. Endothelial receptors involved include intercellular adhesion molecule 1 (ICAM-1) uptake is either via endocytosis or by a unique cellular structure termed an invasome.

Vasculogenesis is induced though an increased production in vascular endothelial growth factor (VEGF) by infected host cells. Currently it is poorly understood how the pathogen provokes overexpression but the mechanism has been determined to be protease sensitive.

Epidemiology. The role of cats and in particular kittens, as reservoirs to *B. henselae* has been firmly established. Infected felines are asymptomatic and show no signs of illness. Cat fleas (*Ctenocephalides felis*) serve as vectors to spread the bacteria among cats and have also been suspected of infecting humans. The main path of transmission to humans however, is through scratches and bites by infected cats. *B. henselae* has also been found in ticks and tick bites prior to contraction of CSD have been reported.

In the United States, 24000 cases of CSD are reported yearly, yielding 2000 hospital admissions with an estimated health care cost of \$12 million. Chil-

2. BIOLOGICAL BACKGROUND

dren are more likely to be affected (80%) and incidence is higher in males (60%). The seasonal pattern (occurrences higher in fall/winter) is attributed to cat mating patterns, as well as pet acquisition fluctuations.

2.2.2 *Brucella Abortus*

The Danish physician David Bang first isolated *Brucella abortus* in 1895 from cyetic cattle tissue, investigating a contagious disease causing abortions in cows. *B. abortus* are small, unflagellated proteobacteria with a cell wall consisting of an outer layer of lipopolysaccharide (LPS) (9 nm) and an inner layer of muramyl mucopeptide complexes (3–5 nm). The Gram-negative coccobacilli appear to have evolved from free-living, soil-dwelling species and are closely related to other human pathogens such as *Bartonella* spp., based on 16S rRNA sequences. *Brucella* species were investigated for possible use as warfare agents in the mid 20th century by several armed forces. ([Atluri et al., 2011](#))

Diseases. Brucellosis is a human disease caused by several pathogenic *Brucella* species, most importantly *B. abortus*, *B. melitensis*, *B. canis* and *B. suis*. Onset may be acute or insidious and due to protean symptoms, diagnosis based on clinical presentation alone is difficult. The febrile disease is generalized and may involve many parts of the body, including nervous, skeletal, gastrointestinal, cardiovascular, respiratory and genitourinary systems. Furthermore, as the bacteria spread to other reservoir hosts via their reproductive systems, persistence of infection is crucial to the pathogen and it comes as no surprise that brucellosis can manifest as a chronic disease in humans too.

Fever is the most consistent sign of *Brucella* infection and depending on what specific organs are affected, further symptoms include asthenia, anorexia, nausea, malaise, arthritis, hepatomegaly, splenomegaly, epididymo-orchitis in males, and pulmonary manifestations such as bronchitis or pneumonia. A rare complication (less than 2%), albeit the most lethal, is infective endocarditis. Invasion of the nervous system occurs in less than 5% of cases and often results in meningitis or meningoencephalitis with good prognosis under antimicrobial treatment.

Pathogenesis. Host entry happens primarily via the digestive system but is also possible through the respiratory tract or skin lesions. On the gastrointestinal route, *Brucella* spp. target Peyer's patches (lymphoid nodules localized towards the end of the small intestine) and must therefore pass through acidic conditions in the stomach. This is facilitated by expression of two ureases capable of hydrolyzing urea and producing a protective bicarbonate buffering system. When entering through the respiratory system,

2.2. Select Microbial Pathogens

B. abortus target alveolar macrophages which serve as access point to the lymphatic system therefore facilitating systematic spread.

In order to persist at systemic sites, both active and passive mechanisms for evading the immune system are in place. LPS of the outer cell wall disguises the bacteria from toll-like receptors(TLRs) and expression of two proteins containing toll-interleukin-1 receptor (TIR) domains actively interferes with TLR signaling.

Uptake by macrophages happens via phagocytosis, which is either triggered by nonopsonized bacteria through a lipid raft mediated mechanism or by opsonisation. Although opsonin marked bacteria are 10-fold more likely to be ingested, the number of pathogens reaching their replicatory niche within the host cell is higher for nonopsonized bacteria. Maturation of early endosomes into lysosomes is important for successful infection as preventing acidification (through addition of bafilomycin A) or fusion with lysosomes (through suppression of the late-endosomal GTPase Rab7) interferes with bacterial replication. Nonopsonized bacteria finally become associated with rough ER and begin replication in ER derived vacuoles within the ER-Golgi intermediate compartment (ERGIC). Blocking the small GTPase Sar1 inhibits intracellular replication by preventing acquisition of coat protein complex II, involved in anterograde ER–Golgi transport (COPII) by ER-exit vesicles and the small GTPase Rab2, involved in ER–cis-Golgi traffic, is required for maximal proliferation.

Despite multiplying intra-cellularly in high numbers, host cells are kept alive and are even able to replicate despite infection. Furthermore *Brucella* species are able to interfere with apoptosis, maintaining their replicatory niche, protected from immune response.

Epidemiology. Preferred natural reservoir species for *B. abortus* are cattle (*Bos taurus* and *Bos indicus*) and almost all parts of the world are affected. The disease exists in both domestic and wild animals and is most prevalent in Mediterranean countries, North Africa, throughout the Middle East, India, Central Asia, as well as South and Central America. Zoonosis most often occurs through ingestion of unpasteurized milk products but airborne transmission is also possible, putting professionals involved in animal husbandry at risk. Vertical transmission among reservoir hosts can occur through lactation and horizontal transmission is facilitated by mating and placental discharge associated with aborted gestation. Human-to-human transmission is rare (but has been suspected to be possible via sexual intercourse), making humans dead-end hosts. As opposed to *Bartonella*, immunodeficient patients do not seem to be especially susceptible towards *Brucella* infections.

Worldwide, an estimated 500000 new cases of brucellosis occur annually, making it one of the most prevalent zoonoses. Although usually susceptible

2. BIOLOGICAL BACKGROUND

to combined antibiotic therapies of at least two agents (usually a tetracycline antibiotic combined with an aminoglycoside or rifampin), untreated brucellosis leads to a high degree of morbidity, leading to being classified a neglected zoonosis by the World Health Organization (WHO).

2.2.3 *Listeria Monocytogenes*

The short, Gram-positive bacilli are non-sporeforming facultative anaerobes, capable of growing in a wide temperature range (0–50°C) and in many different environments. Flagellation is temperature dependent with flagellin being expressed and assembled into peritrichous flagella around 20–25°C but not at 37°C. First described in 1924 by Murray after isolation from lymph glands of diseased laboratory animals, the pathogen was found to also infect humans four years later. For much of the time since, listeriosis was considered a rare zoonotic disease and did not receive much attention. It was not until the 1980s, when several food-borne listeria outbreaks caused a shift in interest towards the pathogen, which has since become a well studied facultatively intracellular parasite ([Farber and Peterkin, 1991](#)).

Diseases. Maternal and neonatal listeriosis accounts for almost half of all infections. Listeriosis in pregnancy typically manifests in bacteraemia and presents as a self-limiting febrile disease with flu-like symptoms. Many cases, however are asymptomatic and the first sign of infection is abortion or neonatal listeriosis. Maternal infection does not necessarily carry over to the fetus, especially if proper chemotherapy is administered. Perinatal incidences are divided into early and late onset (>5 days after parturition), with former cases typically resulting in septicemia and latter cases in meningitis. While in early onset cases the predominant route of transmission is transplacental, the situation is less clear in late onset cases. Both the maternal genital tract during child birth and environmental sources have been implicated. Despite antibiotic treatment, overall mortality rates of 30–40% are typical and prognosis for early onset disease is worse, as it is often associated with preterm birth and advanced stage of infection.

Among adults, most cases of listeriosis occur in T-cell deficient individuals. HIV infection, for example, increases incidence 150–300 fold compared to general population control groups. Predisposing conditions include lymphoreticular neoplasms, deliberate immunosuppression (e.g. antirejection treatment after organ transplants), alcoholism and diabetes mellitus. Despite increased susceptibility caused by immunodeficiency, roughly 30% of all infections affect immunocompetent individuals. In healthy adults, consumption of food contaminated with *L. Monocytogenes* can either lead to self-limiting febrile gastroenteritis with short incubation time (<24 h) or invasive listeriosis with much longer incubation periods (3–4 weeks). The systemic

2.2. Select Microbial Pathogens

form of infection often manifests as bacteraemia or as a neurological infection, but can also involve endocarditis and spread to other parts of the body. Central nervous system involvement occurs in as much as 75% of cases and either presents as meningitis or encephalitis. Mortality rates of 35–45% have been reported for listeriosis in adults.

Pathogenesis. The predominant entry path for *L. Monocytogenes* into the human body is via the gastrointestinal tract, where Peyer's patches are targeted. The bacteria can induce cellular uptake (see section 2.1.2), even by non-phagocytic host cells via expression of cell-surface associated interanlins. Upon cell entry, the phagosomal membrane is lysed, mediated by haemolysin secretion, and actin polymerization (see section 2.1.3) provides means of intracellular movement. Listerial haemolysin (listeriolysin) is acid activated with an optimum around pH 5.5, which is reached in late endosomes and membrane lysis is caused by the formation of large transmembrane pores. Growth and replication occur in the cytoplasm and adjacent cells can be entered by pushing against the plasma membrane and forming a pseudopod-like structure which in turn is taken up the neighbor. The resulting double-membrane vacuole can be escaped by cytolysis.

In an extracellular setting, haemolysins serve to rupture erythrocytes in order to generate free iron, a limiting growth factor. Furthermore the non-specific immune system has to be evaded and haemolysins have also been shown to be cytotoxic towards leukocytes. Additionally, expression of bacterial superoxide dismutase mitigates the effect of free superoxide radicals which play an important role in killing phagocytized bacteria.

Epidemiology. Incidence of listeriosis has initially been increasing since its recognition as food-borne disease but effects of awareness and diagnostic methods are unclear. While typically long incubation periods do pose difficulties for clinical diagnosis, the number of susceptible individuals is on the rise and certain aspects modern processing and handling of foods may be beneficial for growth of *L. Monocytogenes*. Disease rates of 2–15 cases per million population have been reported and listeriosis is among the leading cause of lethal food-borne pathogen infections. Most recent data, however suggests, that incidence is decreasing again.

Due to its non-fastidious lifestyle, *L. Monocytogenes* has been isolated from a wide array of ecological niches, including soil, sewage and water (both fresh water bodies and estuaries). High persistence (up to 4 years) in soil samples is problematic when contaminated manure is used as fertilizer and biofilm formation poses challenges for eradication from food processing plants. Additionally, the ability to grow in refrigerated foods and resistance towards heat treatment such as pasteurization warrant alertness and special preventive care. Many different foods have been implicated in listeriosis outbreaks,

2. BIOLOGICAL BACKGROUND

including vegetables (potatoes, radishes and celery), seafood (shrimp, crab-meat and smoked fish), dairy products (soft cheese, pasteurized and unpasteurized milk) and meats (poultry, various types of sausages and pâté).

Despite high prevalence in food (studies have found 20–80% of meat product samples and 1–10% of dairy product samples contaminated with *L. Monocytogenes*), comparatively few successful infections occur. *L. Monocytogenes* is ingested frequently in small doses and stool sample examinations suggest that 10-70% of investigated populations could be intestinal carriers. In immunocompetent individuals, a large number of bacteria (10^6 – 10^9) has to be ingested for contraction with listeriosis to occur.

2.2.4 *Salmonella Enterica*

Diseases.

Pathogenesis.

Epidemiology.

2.2.5 *Shigella Flexneri*

Diseases.

Pathogenesis.

Epidemiology.

2.2.6 *Adenovirus*

Diseases.

Pathogenesis.

Epidemiology.

2.2.7 *Rhinoviruses*

Diseases.

Pathogenesis.

Epidemiology.

2.2.8 Vaccinia Virus

Diseases.

Pathogenesis.

Epidemiology.

2.3 RNA Interference

Chapter 3

Mathematical Background

Modeling the relationship among variables is one of the most important applications of statistical theory. The study of regression analysis (and the closely related notion of correlation) started to form towards the end of the 19th century with Sir Francis Galton's study of height heredity in humans and his observation of regression towards the mean. Over the next few years, Udny Yule and Karl Pearson cast the developed concepts into precise mathematical formulation, in turn building on work performed by Adrien-Marie Legendre and Carl Friedrich Gauss who developed the method of least squares almost a century earlier ([Allen, 1997](#)).

A multiple linear regression model can be written in matrix-vector form as

$$y = X\beta + \varepsilon \quad (3.1)$$

where $y \in \mathbb{R}^n$ is the vector of observations on the dependent variable, the design matrix $X \in \mathbb{R}^{n \times p}$ contains data on the independent variables, $\beta \in \mathbb{R}^p$ is the p -dimensional parameter vector and the error term $\varepsilon \in \mathbb{R}^n$ captures effects not modeled by the regressors. Without loss of generality, all variables are assumed to be expressed as deviations from their means and measured on the same scale.

In order to find unknown coefficients β_i , the ordinary least squares estimator minimizes the residual sum of squares, the squared differences between observed responses and their predictions according to the linear model.

$$\hat{\beta} = \arg \min_{\beta} \|y - X\beta\|^2 \quad (3.2a)$$

$$= (X^T X)^{-1} X^T y \quad (3.2b)$$

Some assumptions are typically associated with linear regression models that yield desirable attributes for the estimates. None of these restrictions are imposed on the explanatory variables; they can be continuous or discrete

3. MATHEMATICAL BACKGROUND

and combined as well as transformed arbitrarily. Furthermore, in practice, it is irrelevant whether the covariates are treated as random variables or as deterministic constants. With exception of the field of econometrics it appears that the majority of literature adheres to the latter interpretation and therefore, statements will not explicitly be conditional on covariate values.

Linearity. The relationship between dependent and independent variables should be linear (after suitable transformations) and individual effects additive. If this cannot be satisfied, a linear model is not suitable.

Full rank. For the matrix $X^T X$ to be invertible, it has to have full rank p . Therefore $n \leq p$ and all covariates must be linearly independent.

Exogeneity. All independent variables should be known exactly i.e. contain no measurement or observation errors as only the mean squared error of the dependent variable is minimized. Additionally, all important causal factors have to be included in the model. Exogeneity implies $E[\varepsilon_i] = 0 \ \forall i$, as well no correlation between regressors and error terms (Hayashi, 2000).

Spherical errors. This includes both homoscedasticity or constant error variance: $E[\varepsilon_i^2] = \sigma^2 \ \forall i$ and uncorrelated errors $E[\varepsilon_i \varepsilon_j] = 0 \ \forall i \neq j$. These two conditions can be written more compactly as $\text{Var}[\varepsilon] = \sigma^2 I_{n \times n}$.

Normality. For the estimated coefficients to gain some additional desirable characteristics, it can be required that the errors ε_i be jointly normally distributed. Together with the above restrictions on expectation and variance, this yields $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_{n \times n})$.

Violations of these assumptions have varying consequences. In case of perfect multicollinearity, the ordinary least squares estimator $\hat{\beta}$ as defined in (3.2b) does not exist. Recovering such a situation is possible by using a generalized matrix inverse (for example the Moore–Penrose pseudoinverse) or employing a regularization scheme such as ridge regression.

Omitting a variable that is both correlated with dependent variables and has an effect on the response (a nonzero true coefficient) will introduce bias in the parameters. The method of instrumental variables can help to produce an unbiased estimator.

The assumption of spherical errors ensures that the least squares estimator is the best linear unbiased estimator in the sense that it has minimal variance among all linear unbiased estimators. Heteroscedasticity and autocorrelation do not cause coefficient estimates to be biased but can introduce bias in OLS estimates of variance, causing inaccurate standard errors. A generalized least squares estimator (for example weighted least squares)

Chapter 4

Writing scientific texts in English

This chapter was originally a separate document written by Reto Spöhel. It is reprinted here so that the template can serve as a quick guide to thesis writing, and to provide some more example material to give you a feeling for good typesetting.

4.1 Basic writing rules

The following rules need little further explanation; they are best understood by looking at the example in the booklet by Knuth et al., §2–§3.

Rule 4.1 Write texts, not chains of formulas.

More specifically, write full sentences that are logically interconnected by phrases like ‘Therefore’, ‘However’, ‘On the other hand’, etc. where appropriate.

Rule 4.2 Displayed formulas should be embedded in your text and punctuated with it.

In other words, your writing should not be divided into ‘text parts’ and ‘formula parts’; instead the formulas should be tied together by your prose such that there is a natural flow to your writing.

4.2 Being nice to the reader

Try to write your text in such a way that a reader enjoys reading it. That’s of course a lofty goal, but nevertheless one you should aspire to!

Rule 4.3 Be nice to the reader.

Give some intuition or easy example for definitions and theorems which might be hard to digest. Remind the reader of notations you introduced

many pages ago – chances are he has forgotten them. Illustrate your writing with diagrams and pictures where this helps the reader. Etc.

Rule 4.4 Organize your writing.

Think carefully about how you subdivide your thesis into chapters, sections, and possibly subsections. Give overviews at the beginning of your thesis and of each chapter, so the reader knows what to expect. In proofs, outline the main ideas before going into technical details. Give the reader the opportunity to ‘catch up with you’ by summing up your findings periodically.

Useful phrases: ‘So far we have shown that ...’, ‘It remains to show that ...’, ‘Recall that we want to prove inequality (7), as this will allow us to deduce that ...’, ‘Thus we can conclude that Next, we would like to find out whether ...’, etc.

Rule 4.5 Don’t say the same thing twice without telling the reader that you are saying it twice.

Repetition of key ideas is important and helpful. However, if you present the same idea, definition or observation twice (in the same or different words) without telling the reader, he will be looking for something new where there is nothing new.

Useful phrases: ‘Recall that [we have seen in Chapter 5 that] ...’, ‘As argued before / in the proof of Lemma 3, ...’, ‘As mentioned in the introduction, ...’, ‘In other words, ...’, etc.

Rule 4.6 Don’t make statements that you will justify later without telling the reader that you will justify them later.

This rule also applies when the justification is coming right in the next sentence! The reasoning should be clear: if you violate it, the reader will lose valuable time trying to figure out on his own what you were going to explain to him anyway.

Useful phrases: ‘Next we argue that ...’, ‘As we shall see, ...’, ‘We will see in the next section that ..., etc.

4.3 A few important grammar rules

Rule 4.7 There is (almost) *never* a comma before ‘that’.

It’s really that simple. Examples:

We assume that ...
Wir nehmen an, dass ...

It follows that ...

Daraus folgt, dass ...

'thrice' is a word that is seldom used.

'thrice' ist ein Wort, das selten verwendet wird.

Exceptions to this rule are rare and usually pretty obvious. For example, you may end up with a comma before 'that' because 'i.e.' is spelled out as 'that is':

For $p(n) = \log n/n$ we have ... However, if we choose p a little bit higher, that is $p(n) = (1 + \varepsilon) \log n/n$ for some $\varepsilon > 0$, we obtain that ...

Or you may get a comma before 'that' because there is some additional information inserted in the middle of your sentence:

Thus we found a number, namely n_0 , that satisfies equation (13).

If the additional information is left out, the sentence has no comma:

Thus we found a number that satisfies equation (13).

(For 'that' as a relative pronoun, see also Rules 4.9 and 4.10 below.)

Rule 4.8 There is usually no comma before 'if'.

Example:

A graph is not 3-colorable if it contains a 4-clique.

Ein Graph ist nicht 3-färbbar, wenn er eine 4-Clique enthält.

However, if the 'if' clause comes first, it is usually separated from the main clause by a comma:

If a graph contains a 4-clique, it is not 3-colorable .

Wenn ein Graph eine 4-Clique enthält, ist er nicht 3-färbbar.

There are more exceptions to these rules than to Rule 4.7, which is why we are not discussing them here. Just keep in mind: don't put a comma before 'if' without good reason.

Rule 4.9 Non-defining relative clauses have commas.

Rule 4.10 Defining relative clauses have no commas.

In English, it is very important to distinguish between two types of relative clauses: defining and non-defining ones. This is a distinction you absolutely need to understand to write scientific texts, because mistakes in this area actually distort the meaning of your text!

It's probably easier to explain first what a *non-defining* relative clause is. A non-defining relative clauses simply gives additional information *that could also be left out* (or given in a separate sentence). For example, the sentence

4. WRITING SCIENTIFIC TEXTS IN ENGLISH

The WeirdSort algorithm, which was found by the famous mathematician John Doe, is theoretically best possible but difficult to implement in practice.

would be fully understandable if the relative clause were left out completely. It could also be rephrased as two separate sentences:

The WeirdSort algorithm is theoretically best possible but difficult to implement in practice. [By the way,] WeirdSort was found by the famous mathematician John Doe.

This is what a non-defining relative clause is. *Non-defining relative clauses are always written with commas.* As a corollary we obtain that you cannot use 'that' in non-defining relative clauses (see Rule 4.7!). It would be wrong to write

The WeirdSort algorithm, that was found by the famous mathematician John Doe, is theoretically best possible but difficult to implement in practice.

A special case that warrants its own example is when 'which' is referring to the entire preceding sentence:

Thus inequality (7) is true, which implies that the Riemann hypothesis holds.

As before, this is a non-defining relative sentence (it could be left out) and therefore needs a comma.

So let's discuss *defining* relative clauses next. A defining relative clause tells the reader *which specific item the main clause is talking about*. Leaving it out either changes the meaning of the sentence or renders it incomprehensible altogether. Consider the following example:

The WeirdSort algorithm is difficult to implement in practice. In contrast, the algorithm that we suggest is very simple.

Here the relative clause 'that we suggest' cannot be left out – the remaining sentence would make no sense since the reader would not know which algorithm it is talking about. This is what a defining relative clause is. *Defining relative clauses are never written with commas.* Usually, you can use both 'that' and 'which' in defining relative clauses, although in many cases 'that' sounds better.

As a final example, consider the following sentence:

For the elements in \mathcal{B} which satisfy property (A), we know that equation (37) holds.

4.4. Things you (usually) don't say in English

Table 4.1: Things you (usually) don't say

It holds (that) ...	We have ...	<i>Es gilt ...</i>
('Equation (5) holds.' is fine, though.)		
x fulfills property \mathcal{P} .	x satisfies property \mathcal{P} .	x erfüllt Eigenschaft \mathcal{P} .
in average	on average	im Durchschnitt
estimation	estimate	Abschätzung
composed number	composite number	zusammengesetzte Zahl
with the help of	using	mit Hilfe von
surely	clearly	sicher, bestimmt
monotonously increasing	monotonically incr.	monoton steigend
(Actually, in most cases 'increasing' is just fine.)		

This sentence does not make a statement about all elements in \mathcal{B} , only about those satisfying property (A). The relative clause is *defining*. (Thus we could also use 'that' in place of 'which'.)

In contrast, if we add a comma the sentence reads

For the elements in \mathcal{B} , which satisfy property (A), we know that equation (37) holds.

Now the relative clause is *non-defining* – it just mentions in passing that all elements in \mathcal{B} satisfy property (A). The main clause states that equation (37) holds for *all* elements in \mathcal{B} . See the difference?

4.4 Things you (usually) don't say in English – and what to say instead

Table 4.1 lists some common mistakes and alternatives. The entries should not be taken as gospel – they don't necessarily mean that a given word or formulation is wrong under all circumstances (obviously, this depends a lot on the context). However, in nine out of ten instances the suggested alternative is the better word to use.

Chapter 5

Typography

5.1 Punctuation

Rule 5.1 Use opening (‘) and closing (’) quotation marks correctly.

In L^AT_EX, the closing quotation mark is typed like a normal apostrophe, while the opening quotation mark is typed using the French *accent grave* on your keyboard (the *accent grave* is the one going down, as in *frère*).

Note that any punctuation that *semantically* follows quoted speech goes inside the quotes in American English, but outside in Britain. Also, Americans use double quotes first. Oppose

“Using ‘lasers,’ we punch a hole in … the Ozone Layer,” Dr. Evil said.

to

‘Using “lasers”, we punch a hole in … the Ozone Layer’, Dr. Evil said.

Rule 5.2 Use hyphens (-), en-dashes (–) and em-dashes (—) correctly.

A hyphen is only used in words like ‘well-known’, ‘3-colorable’ etc., or to separate words that continue in the next line (which is known as hyphenation). It is entered as a single ASCII hyphen character (-).

To denote ranges of numbers, chapters, etc., use an en-dash (entered as two ASCII hyphens --) with no spaces on either side. For example, using Equations (1)–(3), we see…

As the equivalent of the German *Gedankenstrich*, use an en-dash with spaces on both sides – in the title of Section 4.4, it would be wrong to use a hyphen instead of the dash. (Some English authors use the even longer emdash (—)

instead, which is typed as three subsequent hyphens in L^AT_EX. This emdash is used without spaces around it—like so.)

5.2 Spacing

Rule 5.3 Do not add spacing manually.

You should never use the commands `\v` (except within tabulars and arrays), `\u` (except to prevent a sentence-ending space after Dr. and such), `\vspace`, `\hspace`, etc. The choices programmed into L^AT_EX and this style should cover almost all cases. Doing it manually quickly leads to inconsistent spacing, which looks terrible. Note that this list of commands is by no means conclusive.

Rule 5.4 Judiciously insert spacing in maths where it helps.

This directly contradicts Rule 5.3, but in some cases T_EX fails to correctly decide how much spacing is required. For example, consider

$$f(a, b) = f(a + b, a - b).$$

In such cases, inserting a thin math space `\,` greatly increases readability:

$$f(a, b) = f(a + b, \, a - b).$$

Along similar lines, there are variations of some symbols with different spacing. For example, Lagrange's Theorem states that $|G| = [G : H]|H|$, but the proof uses a bijection $f: aH \rightarrow bH$. (Note how the first colon is symmetrically spaced, but the second is not.)

Rule 5.5 Learn when to use `\u` and `\o`.

Unless you use ‘french spacing’, the space at the end of a sentence is slightly larger than the normal interword space.

The rule used by T_EX is that any space following a period, exclamation mark or question mark is sentence-ending, except for periods preceded by an upper-case letter. Inserting `\,` before a space turns it into an interword space, and inserting `\o` before a period makes it sentence-ending. This means you should write

1 Prof.\ Dr.\ A. Steger is a member of CADMO\@.
2 If you want to write a thesis with her, you
3 should use this template.

which turns into

Prof. Dr. A. Steger is a member of CADMO. If you want to write a thesis with her, you should use this template.

The effect becomes more dramatic in lines that are stretched slightly during justification:

Prof. Dr. A. Steger is a member of CADMO. If you

Rule 5.6 Place a non-breaking space (~) right before references.

This is actually a slight simplification of the real rule, which should invoke common sense. Place non-breaking spaces where a line break would look ‘funny’ because it occurs right in the middle of a construction, especially between a reference type (Chapter) and its number.

5.3 Choice of ‘fonts’

Professional typography distinguishes many font attributes, such as family, size, shape, and weight. The choice for sectional divisions and layout elements has been made, but you will still occasionally want to switch to something else to get the reader’s attention. The most important rule is very simple.

Rule 5.7 When emphasising a short bit of text, use `\emph`.

In particular, *never* use bold text (`\textbf`). Italics (or Roman type if used within italics) avoids distracting the eye with the huge blobs of ink in the middle of the text that bold text so quickly introduces.

Occasionally you will need more notation, for example, a consistent typeface used to identify algorithms.

Rule 5.8 Vary one attribute at a time.

For example, for WEIRDSORT we only changed the shape to small caps. Changing two attributes, say, to bold small caps would be excessive (L^AT_EX does not even have this particular variation). The same holds for mathematical notation: the reader can easily distinguish g_n , $G(x)$, \mathcal{G} and G .

Rule 5.9 Never underline or uppercase.

No exceptions to this one, unless you are writing your thesis on a typewriter. Manually. Uphill both ways. In a blizzard.

5.4 Displayed equations

Rule 5.10 Insert paragraph breaks *after* displays only where they belong. Never insert paragraph breaks *before* displays.

5. TYPOGRAPHY

\LaTeX translates sequences of more than one linebreak (i.e., what looks like an empty line in the source code) into a paragraph break in almost all contexts. This also happens before and after displays, where extra spacing is inserted to give a visual indication of the structure. Adding a blank line in these places may look nice in the sources, but compare the resulting display

$$a = b$$

to the following:

$$a = b$$

The first display is surrounded by blank lines, but the second is not. It is bad style to start a paragraph with a display (you should always tell the reader what the display means first), so the rule follows.

Rule 5.11 Never use `eqarray`.

It is at the root of most ill-spaced multiline displays. The *amsmath* package provides better alternatives, such as the `align` family

$$\begin{aligned} f(x) &= \sin x, \\ g(x) &= \cos x, \end{aligned}$$

and `multline` which copes with excessively long equations:

$$\begin{aligned} \mathbb{P}[X_{t_0} \in (z_0, z_0 + dz_0], \dots, X_{t_n} \in (z_n, z_n + dz_n)] \\ = v(dz_0) K_{t_1}(z_0, dz_1) K_{t_2 - t_1}(z_1, dz_2) \cdots K_{t_n - t_{n-1}}(z_{n-1}, dz_n). \end{aligned}$$

5.5 Floats

By default this style provides floating environments for tables and figures. The general structure should be as follows:

```
1 \begin{figure}
2   \centering
3   % content goes here
4   \caption{A short caption}
5   \label{some-short-label}
6 \end{figure}
```

Note that the label must follow the caption, otherwise the label will refer to the surrounding section instead. Also note that figures should be captioned at the bottom, and tables at the top.

The whole point of floats is that they, well, *float* to a place where they fit without interrupting the text body. This is a frequent source of confusion and changes; please leave it as is.

Rule 5.12 Do not restrict float movement to only ‘here’ (`h`).

If you are still tempted, you should avoid the float altogether and just show the figure or table inline, similar to a displayed equation.

Chapter 6

First Chapter SfS Template

6.1 To include a picture

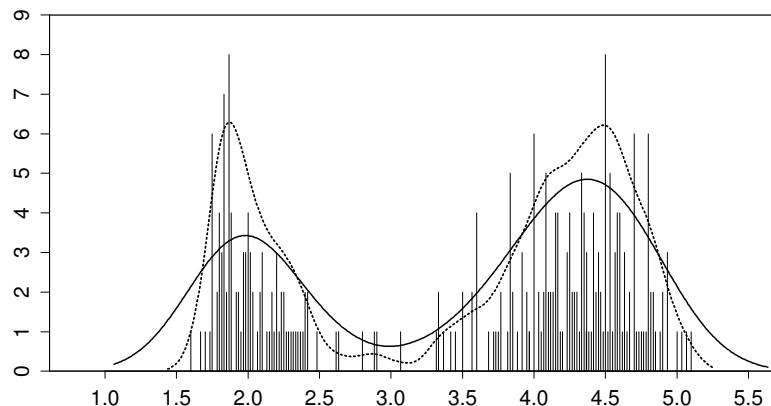


Figure 6.1: Old Faithful Geyser eruption lengths, $n = 272$; binned data and two (Gaussian) kernel density estimates ($\times 10$) with $h = h^* = .3348$ and $h = .1$ (dotted).

Or also with `includegraphics`:

6.2 To make a proof

Proof $1 + 1 = 2$

□

6.3 To include R code

See information in Appendix [A](#).

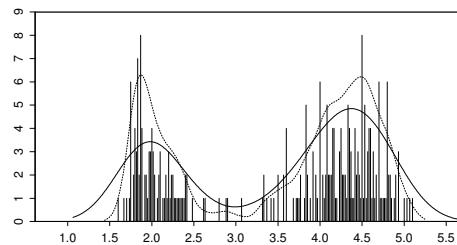


Figure 6.2: Old Faithful Geyser eruption lengths, $n = 272$; binned data and two (Gaussian) kernel density estimates ($\times 10$) with $h = h^* = .3348$ and $h = .1$ (dotted).

6.4 Other information

Put a text between quotes: make sure to use nice quotes, such as “quote”.

Cite a document in the bibliography (an example here): [Gelman et al. \(2008\)](#). Or mention that [Konis](#) (a person) or [Hastie et al.](#) (multiple persons) have already done quite a bit work.

Referencing a different part of your work: please refer to Appendix [A](#).

Appendix A

Complementary information

Additional material. For example long mathematical derivations could be given in the appendix. Or you could include part of your code that is needed in printed form. You can add several Appendices to your thesis (as you can include several chapters in the main part of your work).

A.1 Including R code with verbatim

A simple (rather too simple, see [A.2](#)) way to include code or *R* output is to use `verbatim`. It just prints the text however it is (including all spaces, “strange” symbols,...) in a slightly different font.

```
## loading packages
library(RBGL)
library(Rgraphviz)
library(boot)

## global variables
X_MAX <- 150
```

This allows me to put as many spaces as I want.
I can also use \ and ' and & and all the rest that is usually only accepted in the math mode.

I can also make as
many
line
breaks as
I want... and
where I want.

A.2 Including R code with the *listings* package

However, it is much nicer to use the *listings* package to include R code in your report. It allows you to number the lines, color the comments differently than the code, and so on.

```
1 ## example to generate an .eps file with the function ps.latex()
2 ## Author: Sarah Gerster and Martin Maechler
3 ## Last revision: 16 Aug 2011
4
5 require("sfsmisc") # pdf.latex(), pdf.end(), etc
6
7 pdf.latex(file='test_plot.pdf') #, main=TRUE)
8 ## no main=TRUE is needed to leave enough space for the plot title
9 ## but see below
10
11 ## make sure the legends are large enough
12 par(cex=1.5)
13
14 ## Make sure your lines are "visible" enough. Otherwise your plot
15 ## won't look very nicely in your text.
16 plot(-10:10, (-10:10)**2, type="l", lty=5,
17       xlab="my_x", ylab="my_y",
18       ## no main title: NOT recommended for figures in text which
19       ## have a \caption{...}
20       lwd=4, col='blue')
21 lines(-10:10, 0:20, type="p", lwd=4, pch=23, col='red')
22 legend(-3, 90, c("func1", "func2"), lwd=4, col=c('blue', 'red'),
23         lty=c(1,1), cex=1)
24 pdf.end() # starts the previewer (which refreshes itself;
25           # at least on Linux at SfS
```

A.3 Using Sweave to include R code (and more) in your report

The easiest (and most elegant) way to include R code and its output (and have all your figures up to date with your report) is to use Sweave. You can find an introduction Sweave in /u/sfs/StatSoftDoc/Sweave/Sweave-tutorial.pdf.

Appendix B

Yet another appendix....

B.1 Description

Something details.

Something else other definition.

B.2 Tables

Refer to Table B.1 to see a left justified table with caption on top.

Table B.1: Results.	
Student	Grade
Marie	6
Alain	5.5
Josette	4.5
Pierre	5

Bibliography

- Alberts, B., J. Wilson, and T. Hunt (2008). *Molecular biology of the cell*. New York: Garland Science.
- Allen, M. P. (1997). The origins and uses of regression analysis. In *Understanding Regression Analysis*, pp. 1–5. Springer US.
- Anderson, B. and M. Neuman (1997, apr.). *Bartonella* spp. as emerging human pathogens. *Clin. Microbiol. Rev.* 10(2), 203–219.
- Atluri, V. L., M. N. Xavier, M. F. de Jong, A. B. den Hartigh, and R. M. Tsolis (2011, sep.). Interactions of the Human Pathogenic Brucella Species with Their Hosts. *Annual Review of Microbiology* 65(1), 523–541.
- Baltimore, D. (1971, sep.). Expression of animal virus genomes. *Microbiol. Mol. Biol. Rev.* 35(3), 235–241.
- Craighead, J. E. (2000). *Pathology and Pathogenesis of Human Viral Disease*. San Diego: Academic Press.
- Czyż, D. M., L.-P. Potluri, N. Jain-Gupta, S. P. Riley, J. J. Martinez, T. L. Steck, S. Crosson, H. A. Shuman, and J. E. Gabay (2014, jan.). Host-directed antimicrobial drugs with broad-spectrum efficacy against intracellular bacterial pathogens. *mBio* 5(4), e01534–14.
- Dehio, C. (2005, aug.). Bartonella-host-cell interactions and vascular tumour formation. *Nat Rev Micro* 3(8), 621–631.
- Dobson, A. P. and E. R. Carper (1996, feb.). Infectious Diseases and Human Population History. *BioScience* 46(2), 115–126 CR – Copyright © 1996 Oxford Universi.
- Farber, J. M. and P. I. Peterkin (1991, sep.). Listeria monocytogenes, a food-borne pathogen. *Microbiological Reviews* 55(3), 476–511.

BIBLIOGRAPHY

- Gelman, A., A. Jakulin, M. G. Pittau, and Y.-S. Su (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* 2(4), 1360–1383.
- Haglund, C. M. and M. D. Welch (2011, oct.). Pathogens and polymers: Microbe–host interactions illuminate the cytoskeleton. *The Journal of Cell Biology* 195(1), 7–17.
- Harrison, S. C. (2008, jul.). Viral membrane fusion. *Nat Struct Mol Biol* 15(7), 690–698.
- Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*, Volume 1.
- Hawn, T. R., J. A. Shah, and D. Kalman (2015, mar.). New tricks for old dogs: countering antibiotic resistance in tuberculosis with host-directed therapeutics. *Immunological Reviews* 264(1), 344–362.
- Hayashi, F. (2000). *Econometrics*. Princeton, NJ [u.a.]: Princeton Univ. Press.
- Hulo, C., E. de Castro, P. Masson, L. Bougueret, A. Bairoch, I. Xenarios, and P. Le Mercier (2011, jan.). ViralZone: a knowledge resource to understand virus diversity. *Nucleic acids research* 39(Database issue), D576–82.
- Kahneman, D. (2012, mar.). Quotation: Kahneman on Contingencies. *Journal of the Experimental Analysis of Behavior* 97(2), 182.
- Konis, K. P. (2007). *Linear programming algorithms for detecting separated data in binary logistic regression models*. Doctoral, University of Oxford.
- Kopecko, D. J., L. Hu, and K. J. M. Zaal (2001, aug.). *Campylobacter jejuni* – microtubule-dependent invasion. *Trends in Microbiology* 9(8), 389–396.
- Leirião, P., C. D. Rodrigues, S. S. Albuquerque, and M. M. Mota (2004, dec.). Survival of protozoan intracellular parasites in host cells. *EMBO Reports* 5(12), 1142–1147.
- Littman, R. J. (2009, oct.). The Plague of Athens: Epidemiology and Paleopathology. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine* 76(5), 456–467.
- Prussia, A., P. Thepchatri, J. P. Snyder, and R. K. Plemper (2011). Systematic Approaches towards the Development of Host-Directed Antiviral Therapeutics. *International Journal of Molecular Sciences* 12(6), 4027.
- Rolain, J. M. and D. Raoult (2006). *Bartonella* spp. In *Principles and Practice of Clinical Bacteriology*, pp. 295–304. John Wiley & Sons, Ltd.

Bibliography

- Silver, L. L. (2011, jan.). Challenges of Antibacterial Discovery. *Clinical Microbiology Reviews* 24(1), 71–109.
- Smith, A. (2012). Cellular Biochemistry II: Cellular Infection.
- Taubenberger, J. and D. M. Morens (2006). 1918 Influenza: the Mother of All Pandemics. *Emerging Infectious Diseases Volume* 12(1).
- Treadgold, W. (1997, jun.). *A History of the Byzantine State and Society*. Stanford: Stanford University Press.
- World Health Organization (2003). Fact sheet No. 211: Influenza.
- World Health Organization (2012). Mortality and global health estimates: Causes of death.
- World Health Organization (2014). Antimicrobial resistance: global report on surveillance 2014. Technical report, World Health Organization.
- Zietz, B. P. and H. Dunkelberg (2004). The history of the plague and the research on the causative agent Yersinia pestis. *International Journal of Hygiene and Environmental Health* 207(2), 165–178.

Epilogue

A few final words. Test 2

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):

First name(s):

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Signature(s)

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.