Who’s Next? Identifying Risk Factors for Subjects of Targeted Attack.

Martin Lee

Senior Analyst
Symantec.cloud

Characteristics of Targeted Attacks

"GCHQ now sees real and credible threats to cyber security of an unprecedented scale, diversity and complexity. We've seen determined and successful efforts to:

steal intellectual property;

take commercially sensitive data, such as key negotiating positions;
access government and defence related information;

disrupt government and industry service; and,

exploit information security weaknesses through the targeting of partners, subsidiaries and supply chains at home and abroad."

Iain Lobban, Director GCHQ

Source:
Executive Companion, 10 Steps to Cyber Security. Pub. Cabinet Office (2012)

Characteristics of Targeted Attacks

Targeted	**Non-Targeted**
Attack relevant to interests of recipient	No regard to recipient
Low copy number	High copy number
Bespoke malware	Often kit based
Obscure business model	Clear revenue stream

How Do We Identify Them?

Remove high volume attacks.

Semi-manually analyse remainder:

False positives	Proof of concepts	Targeted attacks
Emailed executables ‘Broken’ documents	Botnet prototypes Script kiddies	Evidence of target selection Sophistication

Context

April 2008 – January 2012:

72500 targeted attack emails.

Sent to 28 300 email addresses.

~500 000 email malware / day.

11 million email addresses.

Annual Targeted Attack Risk

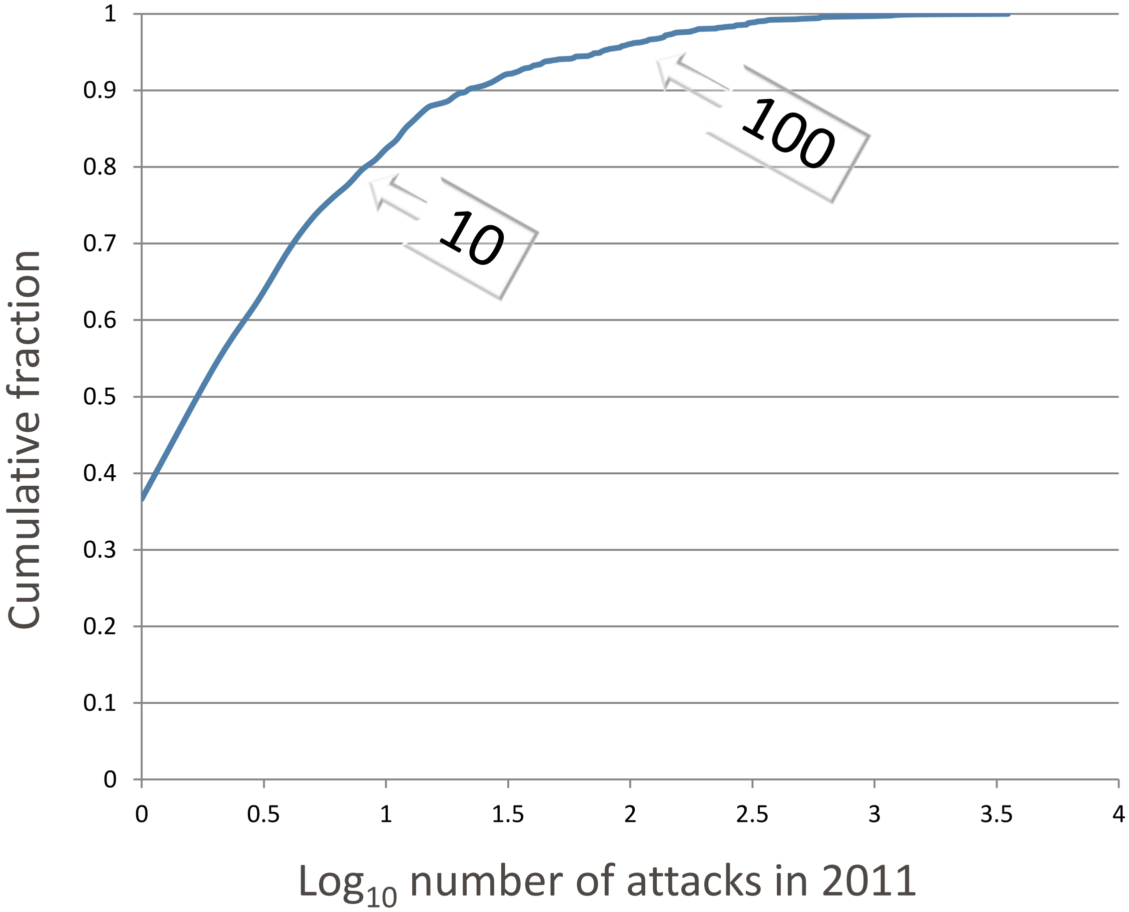
Customers being sent at least 1 targeted attack during 2011:

Type	Ratio Attacked
All Customers	1 : 50.07
SME Customers (<=250 users)	1 : 88.93
Large Customers (>5000 users)	1 : 1.45

Annual office fire risk: 1/588 – 1/161

Source:
Fires in workplace premises: risk data. Holborn et. al.(2002) Fire Safety Journal 37 303-327.

Frequency of attack, 2011



70% received no more than 4.

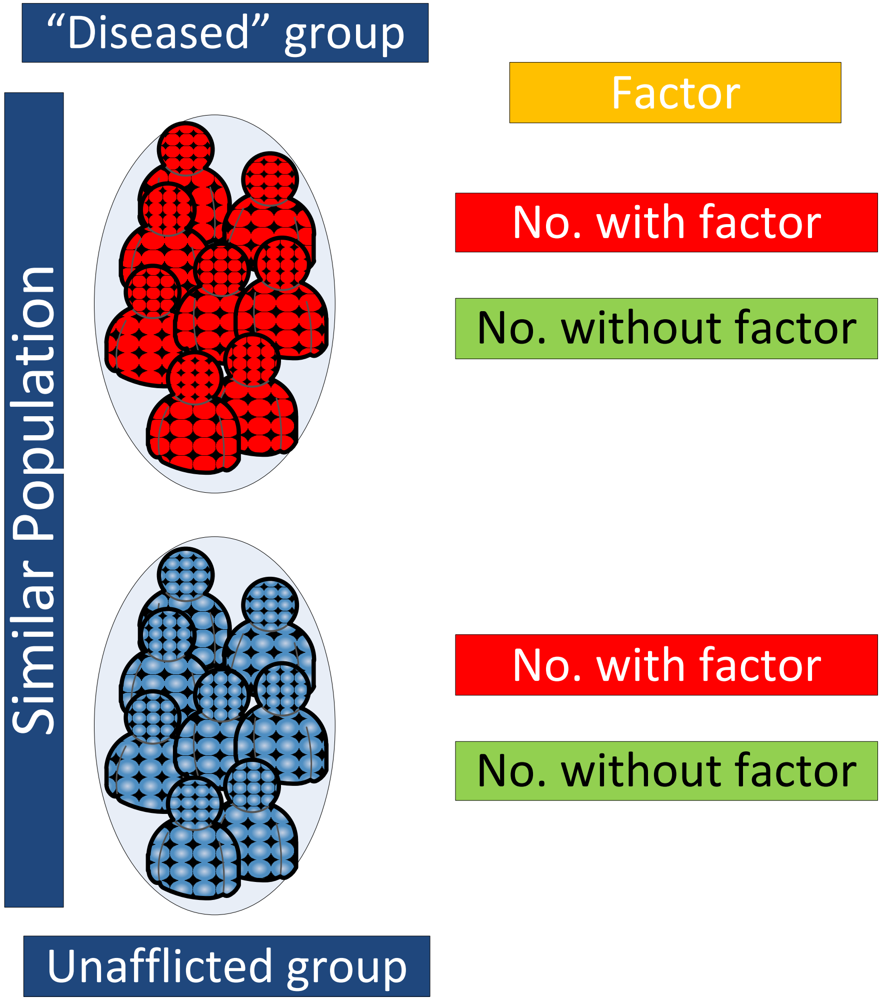
6% received more than 50.

4 receive >1000 attacks.

Building a Risk Based Model

Identifying Risk Factors

Case Control Study



“Diseased” group

Factor

No. with factor

No. without factor

Similar Population

No. with factor

No. without factor

Unafflicted group

Compare likelihood of finding factor in diseased group with that of control group.

Odds Ratio

Calculate strength of association of factor with ‘diseased’ state by comparing probabilities.

	Diseased	Control (unafflicted)
With Risk Factor	p_{11}	p_{10}
Without Risk Factor	p_{01}	p_{00}

$$\text{OR} = \frac{p_{11}\, p_{00}}{p_{10}\, p_{01}}$$

Odds ratio >1 => positive correlation

<1 => negative correlation

Odds Ratio – Standard Error

	Diseased	Control (unafflicted)
With Risk Factor	n_{11}	n_{10}
Without Risk Factor	n_{01}	n_{00}

$$\mathrm{SE}(\log_e \mathrm{OR}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{10}} + \frac{1}{n_{01}} + \frac{1}{n_{00}}}$$

Upper 95% confidence interval = $e^{\log_e \mathrm{OR} + (1.96\ \mathrm{SE}(\log_e \mathrm{OR}))}$

Lower 95% confidence interval = $e^{\log_e \mathrm{OR} - (1.96\ \mathrm{SE}(\log_e \mathrm{OR}))}$

Risk Factors & Protective Factors

	OR	95% CI
Factor 1	x	a - b
Factor 2	y	c - d

Lower 95% CI > 1.0   positive correlation => Risk factor

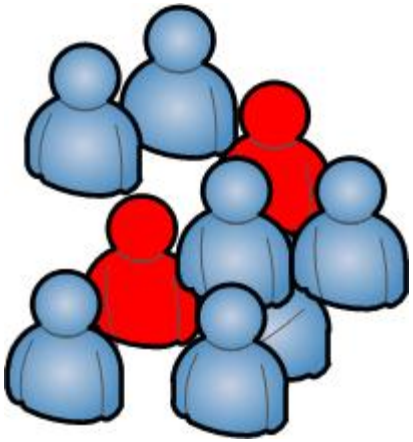
Upper 95% CI < 1.0   negative correlation => Protective factor

Case Control Study Design

Criteria for inclusion in ‘diseased’ and ‘control’ groups.

Match the two groups to minimise differences.

Set of defined factors to test.



Case Control Study Design

“We've seen determined and successful efforts to: steal intellectual property;”

What intellectual property is at risk?

Academic Profile

Dr. *Firstname Surname*

Senior Lecturer in *Subject*

Department of *Subject*

name@university.edu

Recent Publications:

Taxonomy of Higher Education

HESA

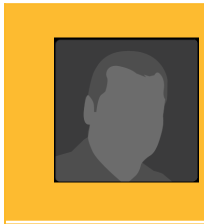
HIGHER EDUCATION STATISTICS AGENCY

UCAS

Joint Academic Coding System (JACS) Version 3.0

	Long Code		Short Code
Computer Science	II00	Computer Sciences	I
Software Engineering	I300		
International Relations	L250	Social Studies	L
War Studies	L252		

Group Classification

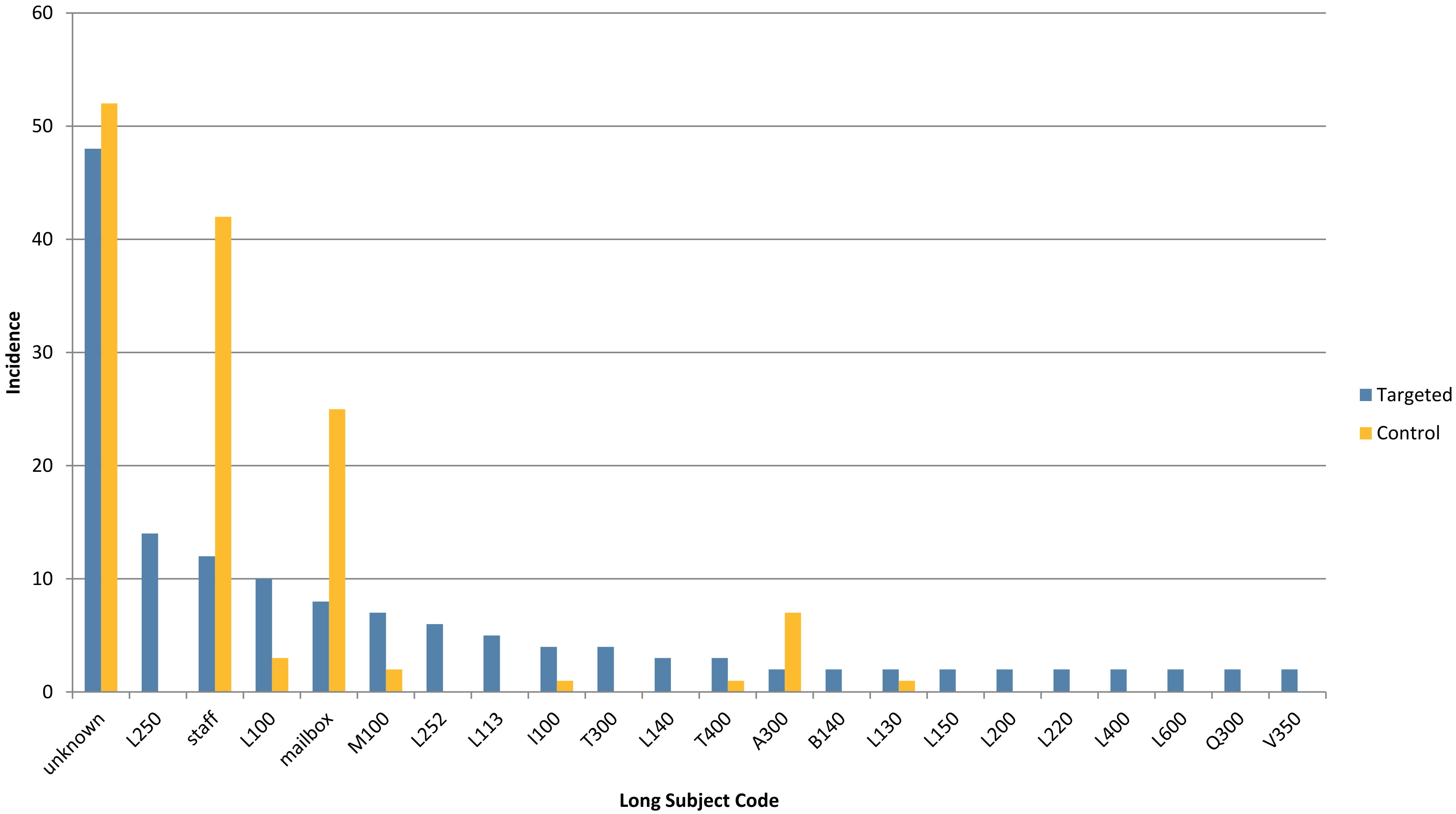
	Received a targeted attack email (n_0) Jan 2010 – Dec 2011	Received a non-targeted attack malware email (n_1)
Classified with subject X	p_{11}	p_{10}
Not classified with subject X	p_{01}	p_{00}

X= JACS3 codes + ‘staff’ + ‘unknown’ + ‘mailbox’

n_0 = 182,

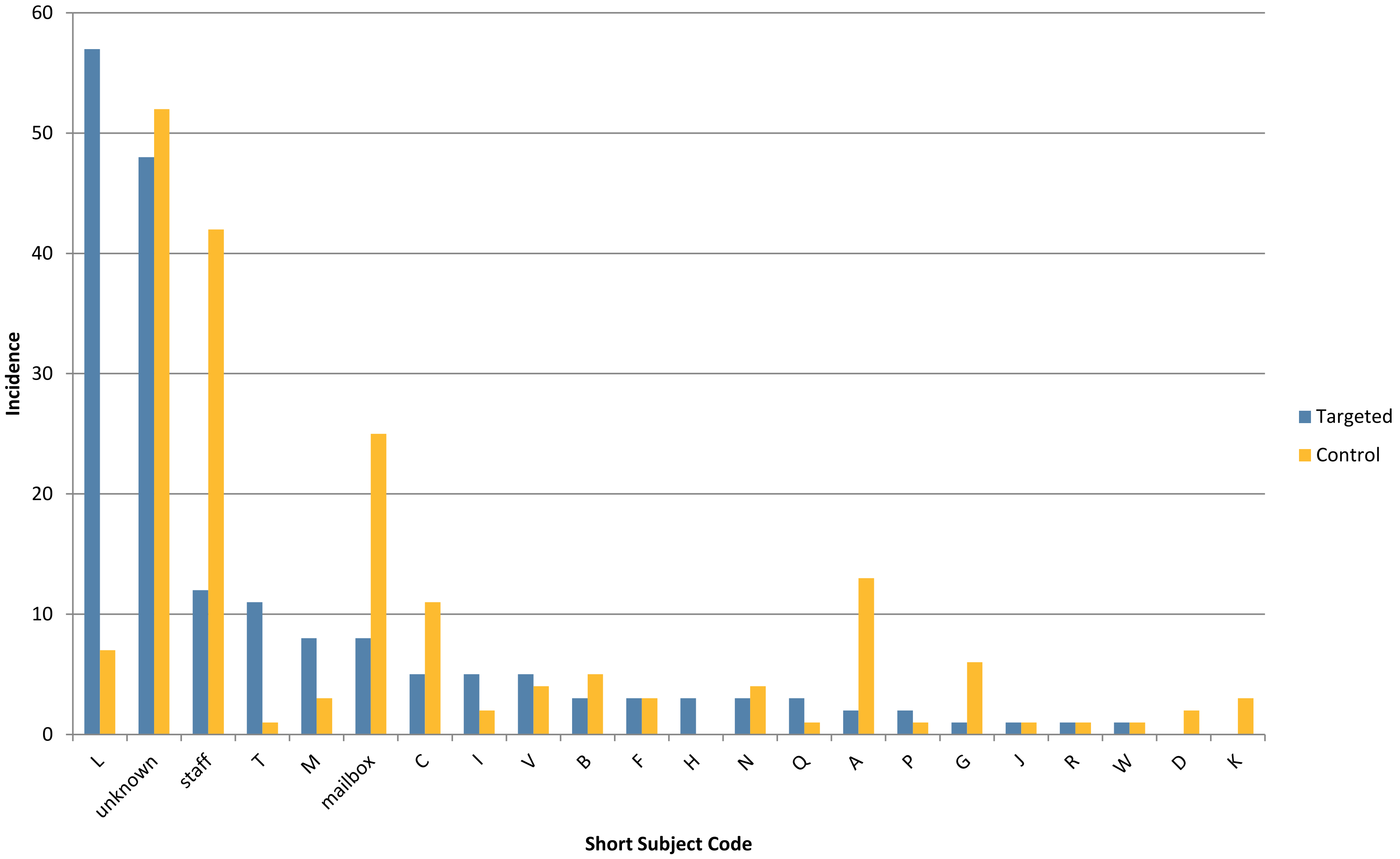
n_1 = 188

Recipient Classification – Long Subject Code



Symantec.cloud

Recipient Classification – Short Subject Code



Results

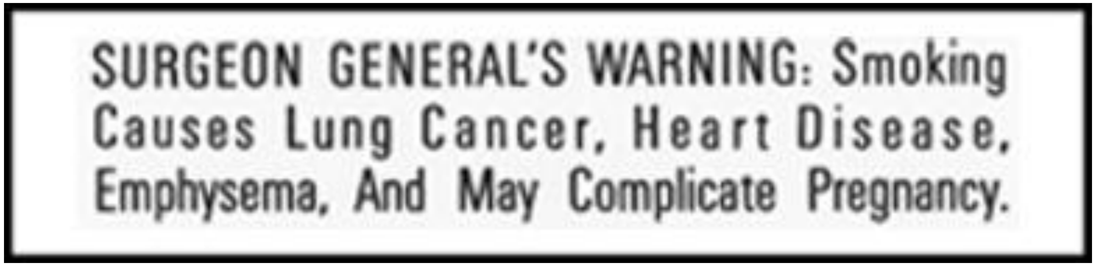
Subject Code	Odds Ratio	95% CI
L (Social Studies)	**11.79**	**(5.21 – 26.70)**
T (Eastern, Asian, African, American, Australasian Studies)	**12.03**	**(1.54 – 94.16)**
I (Computer Sciences)	2.63	(0.50 – 13.72)
G (Mathematical Sciences)	0.17	(0.02 – 1.41)
A (Medicine & Dentistry)	0.15	(0.03 – 0.67)
D (Veterinary Science, Agriculture and Related Subjects)	0	
K (Architecture Building & Planning)	0	
Staff	0.25	(0.12 – 0.48)
Mailbox	0.30	(0.13 – 0.68)

Conclusions

Conclusions

Apply epidemiological analysis to identify those at risk.

Inform those at greatest risk.



SURGEON GENERAL'S WARNING: Smoking Causes Lung Cancer, Heart Disease, Emphysema, And May Complicate Pregnancy.

Enforce policy where most needed.

Symantec.cloud™

Thank you!

Martin Lee

martin_lee@symantec.com

+44 7775 823 278

Thanks: Tony Millington, Prashant Gupta, Steve White, Alistair Johnson, Paul Dominjon.

This document is provided for informational purposes only and is not intended as advertising. All warranties relating to the information in this document, either express or implied, are disclaimed to the maximum extent allowed by law. The information in this document is subject to change without notice.