Scientific Journal of Dermatology and Venereology
Vol. 4, No. 1, 2026, pp. 82–103 · e-ISSN 3025-6208 *journal homepage: phlox.or.id/index.php/sjdv*

Case Study

Reading-Day Yield and Reaction Kinetics of 49 Patch-Test Units in Allergic Contact Cheilitis: An Observational Analytic Case Study in Surakarta

Nila Kusumasari[1], Triasari Oktavriana[2*], Suci Widhiati[1], Wieka Budhiwidayanti[1], Dimas Arief Destiyono[1]

[1]*Department of Dermatology and Venereology, Faculty of Medicine, Universitas Sebelas Maret/Dr. Moewardi General Hospital, Surakarta, Indonesia*

[2]*Department of Dermatology and Venereology, Faculty of Medicine, Universitas Sebelas Maret/Sebelas Maret University Hospital, Surakarta, Indonesia*

ARTICLE INFO

Article history:
Received: July 29, 2026
Revised: August 18, 2026
Accepted: September 7, 2026
Available Online: September 14, 2026
Published: September 15, 2026

Keywords:
Cheilitis
Cosmetics
Dermatitis, allergic contact
Patch tests
Pimecrolimus

***Corresponding author:**
Triasari Oktavriana
E-mail: triasari_oktavriana@staff.uns.ac.id
ORCID: 0000-0002-0879-586X

DOI:
https://doi.org/10.59345/sjdv.v4i1.342

ABSTRACT

Background: Allergic contact cheilitis is a delayed-type hypersensitivity reaction of the vermilion provoked by lip-care and cosmetic products, yet Indonesian data are absent and reading schedules are often truncated.

Objective: To quantify the diagnostic yield each reading day adds, and to describe the reaction kinetics, inter-reading agreement and product-class coherence of the panel.

Methods: An observational analytic case study of a 23-year-old Indonesian woman with seven years of pruritic, dry, fissured lips. The unit of analysis was the patch-test chamber: 49 units, 14 baseline-series allergens and 35 of her own products, graded on the ICDRG scale at 48, 72 and 96 hours (147 observations). Each reading strategy was benchmarked against the union of all three readings, using Fisher's exact and McNemar tests and weighted kappa.

Results: Seventeen units (34.7%, 95% CI 22.9–48.7%) reached a definite positive. A 48 + 72-hour schedule recovered 13 of 17 (76.5%), missing two petrolatum preparations and two lip balms that the day-4 reading recovered (4 gains, 0 losses; exact $p = 0.125$). Six units showed crescendo, eight plateau and six decrescendo kinetics; five of seven doubtful 48-hour reactions became definite by day 4. Weighted agreement fell as the interval widened ($\kappa = 0.881$ to 0.725). Reactivity concentrated in lip-applied products (11 of 12 versus 3 of 23; odds ratio 59.0; $p < 0.001$). Scale cleared by day 15 on pimecrolimus 1% cream with structured avoidance.

Conclusion: Retaining the day-4 reading and testing the patient's own lip products altered the avoidance advice given. No validated severity or quality-of-life instrument was used.

All authors have reviewed and approved the final version of the manuscript.

1. Introduction

Cheilitis is inflammation of the lip, acute or chronic, centred on the vermilion zone and capable of extending to the perioral skin and the oral mucosa.[1-3] It is classified aetiologically into angular, actinic, exfoliative, contact, granulomatous and eczematous forms, and the global prevalence is estimated at 1.3% to 5.8%.[1,2,4] Contact allergy is a substantial and under-recognised share of that burden. In the largest Southeast Asian series, 410 of 5,366 patients referred for patch testing (7.6%) presented with cheilitis and 32% of those were diagnosed with allergic contact cheilitis; the patients' own personal-care products were causative in 73.3%, followed by nickel sulfate in 29.8%, potassium dichromate in 14.5%, castor oil in 14.3% and benzalkonium chloride in 13.0%.[2] The published experience is dominated by lip-applied preparations. Contact cheilitis has been reported to lip balms and lip glosses containing beeswax, other cosmetic waxes or phenoxyethanol, to lipsticks, to toothpaste, to intraoral metal and to foods held against the vermilion, and the demographic signature across those reports is consistent — young women with high cumulative exposure to lip cosmetics.[3,5-11] The allergens implicated in cosmetic contact allergy generally — preservatives, fragrance materials, isothiazolinones and propolis among them — are the same substances declared on the labels of the lip products those patients use.[12-16] For

Indonesia no population estimate exists. National epidemiological data on cheilitis have not been published, and contact-allergy prevalence in the general Indonesian population is unmeasured, so the clinical burden must be inferred from referral practice rather than from surveillance.

The mechanism is type IV, delayed-type hypersensitivity. Sensitisation begins when a hapten of molecular weight below approximately 500 daltons penetrates the stratum corneum, conjugates to carrier protein and is presented by Langerhans and dermal dendritic cells on major histocompatibility complex class I or II molecules; non-specific danger signalling licenses dendritic-cell migration to the draining node and drives clonal expansion of hapten-specific effector and memory T cells.[1,17] Elicitation follows re-exposure in the sensitised individual: cellular stress and innate signalling through toll-like and nucleotide-binding oligomerisation domain-like receptors recruit neutrophils and T cells, antigen-specific effector cells engage antigen-presenting cells, and released interleukin-2, -4 and -10 and interferon-γ amplify keratinocyte and macrophage activation to produce spongiotic dermatitis.[1,17,18] Two properties of that cascade shape the present analysis. The response unfolds over days rather than minutes, which is why patch-test reactions are read at 48 hours and later; and the vermilion is a uniquely permeable site — thin, keratin-poor, without sebaceous protection, continually wetted and mechanically abraded — so a product concentration tolerated elsewhere may elicit at the lip.[1,4,19,20]

Patch testing is the reference standard for confirming contact allergy and the International Contact Dermatitis Research Group scale is its instrument: negative (−), doubtful (+/−), weak positive with erythema, mild oedema and non-vesicular papules (+), strong positive with oedema and vesiculation (++), and extreme positive with bullae (+++).[21] Contemporary practice is explicit that one reading is not enough: a day-2 reading must be supplemented by at least one later reading, because allergic reactions characteristically crescendo while irritant reactions decrescendo, and a single time point cannot distinguish them.[1,22,23] What a truncated schedule costs has been quantified repeatedly: late readings add positives that a day-2 reading alone would miss, whereas omitting the day-2 reading altogether has been reported not to change the final diagnosis, so the two readings are not interchangeable.[24,25] Doubtful reactions, long treated as noise, are now known to declare themselves in a substantial minority of instances.[26,27] Testing the patient's own products is an established adjunct that raises yield beyond the baseline series, with the standing caveats that leave-on preparations are safe to test under occlusion whereas rinse-off products carry an irritant risk, and that a negative result on a finished product does not exclude allergy to a constituent present at low concentration.[28] Quality of life in allergic contact dermatitis is measurably impaired, and it improves once patch testing changes what a patient applies — which is precisely why an instrument should be administered whenever an outcome is to be reported.[29,30] What the literature does not offer is a validated severity index specific to cheilitis; severity is conventionally recorded morphologically, and quality-of-life measurement relies on generic dermatological instruments.

The Department of Dermatology and Venereology of Sebelas Maret University, serving Dr. Moewardi General Hospital and Sebelas Maret University Hospital in Surakarta, Central Java, operates one of the few patch-test services in the region and receives referrals from private clinics and secondary hospitals across a wide catchment. Management follows the national clinical practice guidance of the Indonesian Society of Dermatology and Venereology (PERDOSKI), which is a national consensus document without a digital object identifier and is therefore referred to here by name rather than by a numbered citation: identify the allergen, eliminate exposure, treat the inflammation topically, and educate.[1,31] Two features of that practice setting are analytically important. Indonesian patients present with large personal inventories of locally manufactured cosmetics whose ingredient declarations are not represented in the European baseline series that defines international surveillance, so the incremental yield of personal-product testing is plausibly higher here than the international literature implies.[32-36] Asian and Mediterranean series bear that out: allergen profiles reported from Korea, Taiwan, India, Thailand and Turkey differ from one another and from the European baseline in both the ranking and the frequency of the commonest sensitisers, and the profile shifts again with age.[18,19,37-41] And the day-4 visit is the reading most often sacrificed when clinics are crowded or patients travel far, yet what that omission costs has not been quantified in an Indonesian setting.

Systematic analytic documentation of allergic contact cheilitis using the full three-reading schedule and the patient's own product inventory therefore remains limited at Indonesian tertiary centres. This study analysed a single index case of allergic contact cheilitis at Dr. Moewardi General Hospital and Sebelas Maret University Hospital, Surakarta, taking the patch-test unit rather than the patient as the unit of analysis, in order to quantify the diagnostic yield of each reading strategy against the three-reading union, to classify the reaction kinetics of every reacting unit, to measure agreement between readings, and to test whether reactivity concentrated in lip-applied products.

2. Methods

2.1. Study design and unit of analysis

An observational analytic case study was conducted. One index patient contributed 49 patch-test units — 14 baseline-series allergens and 35 of her own skin-care and cosmetic products — each graded at three reading times, yielding 147 graded observations. The unit of analysis was the patch-test unit, not the patient. Four comparisons were specified before analysis: the diagnostic yield of every reading-day strategy against the union of all three readings; the kinetic pattern of each reacting unit; agreement between reading days;

and whether reactivity concentrated in lip-applied products.

The 49 units derive from one immune system, one anatomical test field and one occlusion period, and are therefore not independent observations. Every interval reported below describes the precision with which this panel is summarised. None is a population estimate, and no result is generalised beyond this patient without explicit statement.

2.2. Ethical approval

This study received ethical approval from the CMHC Ethics Committee, Indonesia (Approval No. 2026/325). Written informed consent was obtained from the patient for participation and for the publication of her clinical photographs and accompanying data. The study was conducted in accordance with the Declaration of Helsinki.

2.3. Setting and participant

The study was conducted at the Department of Dermatology and Venereology, Dr. Moewardi General Hospital and Sebelas Maret University Hospital, Surakarta, Central Java, Indonesia. The patient was a 23-year-old woman referred from a dermatology clinic for allergy testing after a seven-year history of pruritus, dryness and fissuring of the lips. The patient had taken no corticosteroid or antihistamine for two weeks before testing.

2.4. Clinical assessment

History taking covered symptom onset and course, the temporal relation of symptoms to specific products, previous treatment and response, personal and family history of atopy, drug allergy, food allergy and skin disease, and comorbidity. Physical examination recorded general condition, level of consciousness, vital signs, weight, height and body mass index. Dermatological examination recorded the morphology and distribution of lip lesions. Full blood count with differential and a peripheral blood smear were obtained. Clinical photographs of the lips were taken at presentation and at each follow-up visit.

2.5. Patch testing

Testing used the Finn Chamber system with hypoallergenic tape, a permanent marker and cotton swabs moistened with 70% alcohol. Fourteen baseline-series allergens and all 35 products the patient brought were applied. Chambers were placed on the upper back. The patient was required to have taken no corticosteroid or antihistamine for two weeks beforehand, and was instructed not to wet the back, not to undertake activity provoking sweating, not to scratch, and not to lie or lean on her back. Readings were performed at 48 hours (day 2), 72 hours (day 3) and 96 hours (day 4). At the first reading the chambers were removed gently and the reaction was graded after a 20-minute interval to allow resolution of pressure effects; that interval is also the window in which an immediate, non-delayed reaction would be apparent, although the record does not state that any was sought.[42] Grading followed the International Contact Dermatitis Research Group convention. In its standard formulation this is: negative (−); doubtful (+/−), faint macular erythema only; weak positive (+), erythema with infiltration and possibly discrete papules; strong positive (++), erythema, infiltration, papules and vesicles; and extreme positive (+++), intense erythema with coalescing vesicles, bullae or ulceration. Two further codes exist and were not used here: IR for an irritant reaction and NT for a substance not tested.[21,22] The source record's own legend renders the scale slightly more narrowly — (+) as erythema, mild oedema and non-vesicular papules, (++) as oedema with vesicle formation, and (+++) as bulla formation — and the grades analysed here are the grades the record assigned under that legend. The discrepancy between the record's legend and the standard formulation is reported rather than harmonised, because a reader comparing these grades with another series needs to know which wording generated them.

The baseline panel comprised 14 allergens. It is a truncated series rather than the European or North American baseline series, and the omissions matter for this particular case: it contains no fragrance mix I or II, no lanolin alcohol or Amerchol L101, no propylene glycol, no paraben mix, no isothiazolinone preservative, no modern ultraviolet filter other than benzophenone, and none of the lip-cosmetic-specific allergens such as castor oil, shellac, propolis, beeswax, sesame oil or the azo colourants.[21,22,27,32,35] Several of those absent substances are declared in the products that reacted, so the baseline arm of this panel could not have identified them even had the patient been sensitised to them, and the manuscript draws no inference from their absence. The point is not hypothetical. Testing fragrance markers alongside the individual constituents of the fragrance mixes identifies sensitised patients whom the markers alone would miss, and the reported prevalence of fragrance contact allergy varies severalfold between populations.[43-45] Shellac, a colourant binder used in lip cosmetics, is likewise absent from this panel and has been proposed as a marker of fragrance sensitisation in its own right.[46]

2.6. Investigator-applied classifications

Three classifications used in the analysis were applied by the present authors and do not appear in the clinical record; each is stated so that a reader may re-derive or reject it. First, a unit was classified as **lip-applied** if its label named a lip product or if the case history recorded application to the lips; under this primary rule the two petrolatum preparations were lip-applied, because the history records that the patient applied petrolatum to her lips. A pre-specified sensitivity analysis (S1) reclassified both as general emollients. Second, a personal product was classified as **short-contact** if it is removed from the skin within minutes of application — facial washes, toothpastes, exfoliators and micellar waters — and **leave-on** otherwise. Third, **clinical relevance** was classified post hoc using a COADEX-style scheme, since the treating team recorded no relevance assessment. Four categories were used: *current* for a product the patient

was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal; *current (regional)* for a facial product she was using at presentation but applying to the face rather than to the lip, where transfer to the vermilion is plausible but is not documented; *exposed* where a route of exposure to a chemically related substance is documented but the tested substance itself is declared in nothing she used; and *unknown* where no source of exposure is documented at all. The distinction between the first two categories is deliberate: three of the reacting products are facial preparations, and calling them relevant to a lip disease would assert a transfer route the record does not establish.[21,22]

A two-item ordinal index was derived from the morphological descriptors in the record to summarise the clinical course (erythema absent 0 or present 1; scale absent 0, reduced 1, present 2). It is an investigator-constructed summary of what was written down, not an instrument administered to the patient, and it is reported as such throughout.

2.7. Treatment and follow-up

Non-pharmacological management consisted of structured education to avoid the products identified as reactive, to avoid lip licking and wetting the lips with saliva, and to maintain lip hydration. Pharmacological treatment was pimecrolimus 1% cream (Elidel) applied twice daily to the lips, selected as second-line therapy after failure of topical triamcinolone.[47-50] Review visits with photographic documentation took place on days 4, 8 and 15 after treatment began.

2.8. Outcome measures and statistical analysis

The primary outcome was the recovery achieved by each reading strategy, defined as the number of units reaching a definite positive (ICDRG + or ++) under that strategy as a proportion of the 17 units reaching a definite positive at one or more of the three readings. The word recovery is used in preference to sensitivity because the reference set is not an external truth standard: it is the union of the same three readings, so every strategy is evaluated against a target that partly contains it, and the full three-reading schedule attains 100% by construction rather than by performance. The comparison that carries meaning is between the truncated strategies, and the identity of the units each one omits. Secondary outcomes were the kinetic classification of each reacting unit, agreement between reading days, the distribution of reactivity by product class, and the clinical course of the lip lesion.

A unit was classified as **crescendo** when its grade at 96 hours exceeded its grade at 48 hours, **decrescendo** when it was lower, **plateau** when all three readings were identical, and **fluctuating** otherwise. Proportions are reported with Wilson score 95% confidence intervals. Differences between proportions are reported with Newcombe hybrid-score intervals and Cohen's h. Two-by-two associations were tested with Fisher's exact test and summarised by the conditional maximum-likelihood odds ratio with its exact 95% interval, which is the appropriate estimator when a cell contains a single observation.

Two distinct comparisons of definite-positive counts were made and are reported separately, because conflating them would attach the wrong p value to the paper's central claim. The first compares two single readings with each other (72 versus 96 hours, and 48 versus 96 hours). The second compares two reading *strategies* — a 48 + 72-hour schedule against the full three-reading schedule — and it is this second comparison that corresponds to the clinical question of whether the day-4 visit changes the result. Both were tested by an exact binomial (McNemar) test on the discordant units.

Agreement between reading days across the four-level ordinal scale was quantified by linear- and quadratically-weighted Cohen's κ with uncorrected percentile bootstrap 95% intervals from 10,000 resamples, interpreted by the Landis and Koch bands. These coefficients measure agreement between two readings of the same chambers on different days, not agreement between observers; no inter-observer study was performed, and the record does not state how many observers graded the chambers. The resampling is over units, which are not independent, and 36.3% of the resamples contain no strong-positive unit at all, so the intervals are descriptive rather than inferential.

The three quantities that carry the paper's conclusions are defined here explicitly so that a reader can reproduce them. Every proportion is reported with a Wilson score interval, in which for k events in n units and z = 1.96 the limits are $[\hat{p} + z^2/2n \pm z\sqrt{(\hat{p}(1-\hat{p})/n + z^2/4n^2)}] / (1 + z^2/n)$, with $\hat{p} = k/n$. The effect size accompanying every comparison of two proportions is Cohen's $h = 2\cdot\arcsin\sqrt{p_1} - 2\cdot\arcsin\sqrt{p_2}$, which is scale-free and does not degenerate when a proportion approaches zero or one. Agreement between two readings is the weighted kappa $\kappa_w = 1 - (\Sigma_{ij}\, w_{ij}\, O_{ij}) / (\Sigma_{ij}\, w_{ij}\, E_{ij})$, where O_{ij} and E_{ij} are the observed and chance-expected proportions in cell (i, j) of the 4 × 4 grade table and the weights are $w_{ij} = |i - j| / (k - 1)$ for the linear form and $w_{ij} = [(i - j) / (k - 1)]^2$ for the quadratic form, with k = 4 grades.

Because the investigator-applied product classifications could reasonably have been drawn differently, the principal association was recomputed under ten alternative classification rules, and the range of odds ratios and p values across them is reported. Where a vehicle effect and a site effect could be confounded, the vehicle comparison was repeated within strata of site. Six hypothesis tests are reported in total; Holm-adjusted p values are given alongside the unadjusted values, and the adjustment is applied to the interpretation, not merely displayed.

Exact p values are given to three decimal places. Analysis used Python 3.11 with NumPy 2.3 and SciPy 1.17; the analysis script and the transcribed reaction matrix are available from the corresponding author. Reporting follows the CARE recommendations for the case component and reports every analysis specified.

2.9. Analyses that were not performed, and why

Eight analyses that a protocol of this kind might be expected to contain were not performed, because the clinical record cannot support them. Each is set out with its reason, as detailed in Table 1, rather than omitted silently, so that a reader can see the boundary of the evidence. In particular, no Fitzpatrick phototype is reported because none was recorded; no Dermatology Life Quality Index, Skindex-29, itch numerical rating scale, body-surface-area figure or validated severity index is reported because none was administered; and no minimal clinically important difference is applied, because an MCID is defined relative to an instrument and no instrument was used.

Table 1. Analyses specified for this design that the clinical record cannot support.

#	Analysis withheld	Reason
R1	Fitzpatrick phototype	No phototype is recorded anywhere in the source document; assigning one from a clinical photograph would be an unvalidated observer judgement presented as a measurement. Not reported.
R2	DLQI, Skindex-29, itch NRS/VAS, SCORAD, EASI, IGA and any body-surface-area figure	None of these instruments was administered, and none can be reconstructed from a morphological descriptor. No quality-of-life or validated severity score, and no minimal-clinically-important-difference comparison, is reported.
R3	Any test of association between a declared ingredient and patch-test outcome	A declared composition was transcribed only for units that reacted: 15 of 49, of which 13 were definite positives and 2 doubtful, and none of the 29 non-reacting personal products has ingredient data. The missingness is informative, so no denominator exists against which enrichment of an ingredient among reactors could be judged, and no sensitivity, specificity or odds ratio is estimable. Ingredient content is reported as frequency among reacting products only.
R4	Wilcoxon signed-rank testing of a pre-post severity score, correlation of severity with quality of life, and any between-case comparison	There is one patient and no repeated numeric score; these procedures have no data to operate on.
R5	Comparison of this patient's within-panel positivity proportion with published between-patient prevalences	The denominators differ in kind (test units within one patient versus patients within a cohort). Published cohort figures are quoted as context only and are never entered into a significance test against this panel.
R6	Population inference from the 95% intervals reported here	The 49 units share one immune system and one test site and are not independent; the intervals describe the precision with which this panel is summarised, nothing more.
R7	Reconstruction of the patch-test technical protocol beyond what is recorded	The record names the Finn Chamber system, hypoallergenic tape, a 70% alcohol swab, the upper back as the test site, the 2-week steroid and antihistamine washout, the patient restrictions, the 20-minute rest before the first reading, and readings at 48, 72 and 96 h. Chamber diameter, vehicle, the concentrations at which the patient's own products were applied, occlusion time and reader identity are not recorded and are not supplied.
R8	CTCAE grading of adverse effects	No adverse effect of pimecrolimus was recorded; an absent event cannot be graded.

3. Results

3.1. Patient, examination and laboratory findings

A 23-year-old woman presented with a seven-year history of pruritus and dryness of the lips accompanied by fissuring. Symptoms had begun after she started using lip balms and lipsticks, worsened with particular products and improved on their withdrawal. Three months before referral she had attended a private hospital with lip pruritus and smarting accompanied by facial swelling after trying a newly purchased lipstick, and was treated with an oral agent she could not name, triamcinolone ointment twice daily to the lips and a baby soap for facial washing; symptoms improved but recurred whenever treatment stopped. One month before referral symptoms worsened after eating corn and mango and after using a lip balm bought at a cosmetics store. One week before referral the symptoms interfered with daily activity, prompting attendance at a dermatology clinic, a compounded preparation she could not name, and referral for allergy testing. She reported no previous episode of similar complaint, no previous skin disease, and no history of hypertension, diabetes mellitus, atopy, drug allergy, food allergy or chronic illness; family history was negative for all of these. Demographic and clinical details are set out in Table 2, and the referral and assessment sequence, with the reaction counts annotated at each reading, is shown in Figure 1.

On examination she appeared mildly unwell with a Glasgow Coma Scale of E4 M6 V5. Blood pressure was 125/88 mmHg, pulse 84 beats per minute, respiratory rate 20 per minute and temperature 36.4 °C. Weight was 65 kg and height 156 cm, giving a body mass index of 26.7 kg/m^2, recorded in the source as obesity. Dermatological examination of the superior and inferior labial regions showed an erythematous patch with overlying thin scale. The differential diagnosis recorded was allergic contact cheilitis, exfoliative cheilitis and atopic cheilitis. The assessment instruments applied at

this visit, and those that were not administered, are listed in full in Table 3.

Full blood count, differential count and peripheral blood smear were reported as within normal limits, with no haematological abnormality on the smear. Two values printed in the record fall marginally outside their own quoted reference ranges — mean corpuscular haemoglobin concentration 32.8 g/dL against 33.0–36.0, and neutrophils 54.20% against 55.00–80.00 — and are reported here as printed, with the laboratory's own conclusion of no abnormality, as detailed in Table 4. The eosinophil fraction was 3.45%, within the quoted range.

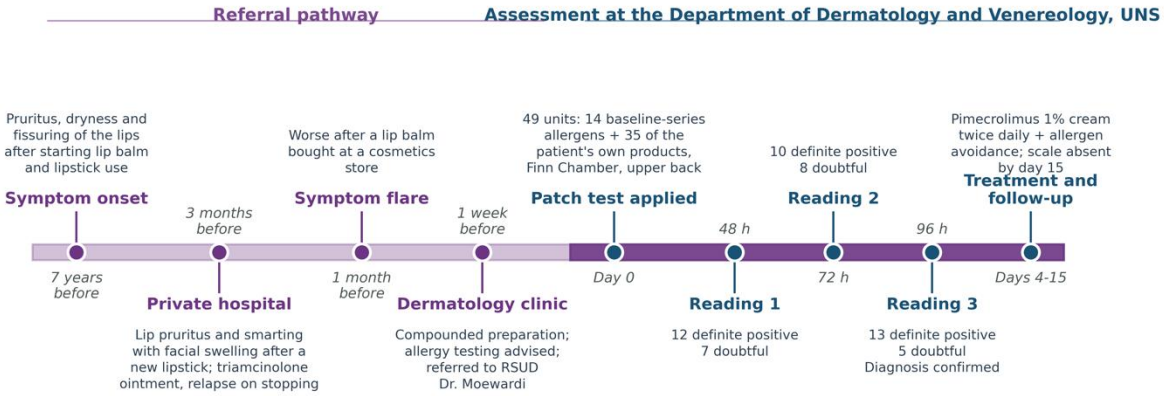

Figure 1. Clinical timeline of allergic contact cheilitis at Dr. Moewardi General Hospital / Sebelas Maret University Hospital, Surakarta. The referral pathway is shown to the left and the assessment at the Department of Dermatology and Venereology to the right, with the number of definite and doubtful reactions annotated at each patch-test reading. Intervals before day 0 are those recorded in the history; calendar dates are not stated in the clinical record.

Table 2. Patient demographic and clinical profile.

Characteristic	Value
Age	23 years
Sex	Female
Presenting complaint	Pruritus, dryness and fissuring of the lips
Duration of symptoms	7 years
Precipitant reported	Onset after starting lip balm and lipstick use; aggravation with specific products, improvement on withdrawal
Fitzpatrick phototype	Not recorded
Ethnicity and occupation	Not recorded
Personal history of atopy	Denied
Personal history of drug or food allergy	Denied
Previous skin disease	Denied
Comorbidity	Hypertension, diabetes mellitus and chronic illness denied
Family history	Similar complaint, skin disease, atopy, drug allergy and food allergy all denied
Previous treatment	Triamcinolone ointment twice daily; petrolatum; an unnamed oral agent; an unnamed compounded preparation
Response to previous treatment	Partial, with relapse whenever treatment was stopped
General condition	Mildly unwell; Glasgow Coma Scale E4 M6 V5
Blood pressure	125/88 mmHg
Pulse	84 beats/min
Respiratory rate	20 breaths/min
Temperature	36.4 °C
Weight, height, body mass index	65 kg, 156 cm, 26.7 kg/m² (recorded as obesity)
Dermatological status at presentation	Regio labialis superior et inferior: erythematous patch with overlying thin scale
Differential diagnosis	Allergic contact cheilitis; exfoliative cheilitis; atopic cheilitis
Final diagnosis	Allergic contact cheilitis due to skin-care and cosmetic products
Setting	Department of Dermatology and Venereology, Faculty of Medicine, Sebelas Maret University / Dr. Moewardi General Hospital, Surakarta, Indonesia

Table 3. Dermatological assessment: instruments applied, findings recorded, and instruments not administered.

Domain	Instrument or method	Recorded result
Lesion morphology	Descriptive dermatological examination	Erythematous patch with overlying thin scale, superior and inferior labial regions
Lesion extent	Body-surface-area or lip-surface-area estimate	**Not administered**
Skin phototype	Fitzpatrick classification	**Not administered**
Contact allergy	Patch test, 49 units, ICDRG grading at 48, 72 and 96 h	17 units definite positive at ≥ 1 reading; see Table 6
Dermoscopy	Dermoscopic examination of the vermilion	**Not performed**
Histopathology	Lip or skin biopsy	**Not performed**
Haematology	Full blood count, differential, peripheral smear	No haematological abnormality (Table 4)
Disease severity	SCORAD, EASI, IGA or any validated index	**Not administered**
Quality of life	DLQI, Skindex-29 or CDLQI	**Not administered**
Symptom intensity	Itch numerical rating scale or visual analogue scale	**Not administered**
Clinical course	Investigator-derived two-item descriptor index (erythema 0–1, scale 0–2)	3 at day 0 and day 4, 2 at day 8, 1 at day 15 (Figure 2B)
Adverse effects	CTCAE grading	**No adverse effect recorded**

Notes: *ICDRG: International Contact Dermatitis Research Group. DLQI: Dermatology Life Quality Index. CDLQI: Children's Dermatology Life Quality Index. SCORAD: SCORing Atopic Dermatitis. EASI: Eczema Area and Severity Index. IGA: Investigator Global Assessment. CTCAE: Common Terminology Criteria for Adverse Events. Entries marked not administered were absent from the clinical record and were not reconstructed.*

Table 4. Laboratory findings.

Investigation	Result	Unit	Reference range	Interpretation
Haemoglobin	14.3	g/dL	12.0-15.6	normal
Haematocrit	44	%	35-45	normal
Leucocytes	8.3	$\times 10^3/\mu L$	4.5-11.0	normal
Platelets	275	$\times 10^3/\mu L$	150-450	normal
Erythrocytes	4.97	$\times 10^6/\mu L$	4.10-5.10	normal
MCV	87.6	fL	80.0-96.0	normal
MCH	28.7	pg	18.0-33.0	normal
MCHC	32.8	g/dL	33.0-36.0	marginally below the quoted range
RDW	13.5	%	11.6-14.6	normal
MPV	11.0	fL	7.2-11.1	normal
Eosinophils	3.45	%	0.00-4.00	normal
Basophils	0.55	%	0.00-2.00	normal
Neutrophils	54.20	%	55.00-80.00	marginally below the quoted range
Lymphocytes	35.10	%	22.00-44.00	normal
Monocytes	6.76	%	0.00-7.00	normal

Notes: *Peripheral blood smear: normochromic normocytic erythrocytes without erythroblasts; leucocyte count within normal limits with atypical lymphocytes and no blasts; platelet count normal with macrothrombocytes, no clumping, even distribution. Conclusion recorded by the laboratory: no haematological abnormality.*

3.2. Distribution of patch-test reactions across the three readings

All 49 units were graded at each of the three readings, giving 147 observations with no missing data. At 48 hours 11 units were weak positive, one was strong positive and 7 were doubtful. At 72 hours the corresponding counts were 9, one and 8, and at 96 hours 12, one and 5; the full distribution of grades at each reading is given in Table 5. The number of definite positives was therefore 12 at 48 hours, 10 at 72 hours and 13 at 96 hours — a non-monotone sequence, falling between the first and second readings and rising again by the third.

Across the three readings combined, 17 of the 49 units (34.7% (95% CI 22.9–48.7%)) reached a definite positive at one or more time point, and 20 units produced any non-negative reaction. Only nickel sulfate 5% reached a strong positive, and it did so at all three readings; no extreme positive reaction and no bulla was recorded. Three of the 17 definite positives were

baseline-series allergens (ethylenediamine 1%, benzophenone 3% and nickel sulfate 5%) and 14 were the patient's own products, so the patient's own inventory accounted for 82.4% of all definite positives.

The complete 49-unit matrix, unit by unit and reading by reading, is presented in Table 6, and the 20 reacting units are displayed grouped by kinetic class in panel A of Figure 2.

Table 5. Distribution of ICDRG reaction grades at each reading (n = 49 units).

Reaction grade	48 h (day 2)	72 h (day 3)	96 h (day 4)
Negative (−)	30	31	31
Doubtful (+/−)	7	8	5
Weak positive (+)	11	9	12
Strong positive (++)	1	1	1
Definite positive (+ or ++)	**12**	**10**	**13**
Any non-negative reaction	**19**	**18**	**18**

Table 6. Complete patch-test result matrix: 49 units read at 48, 72 and 96 hours.

No	Allergen or product	Series	48 h	72 h	96 h
1	Mercaptobenzothiazole 2%	Baseline series	−	−	−
2	Cobalt chloride 1%	Baseline series	−	−	−
3	p-Phenylenediamine 0.1%	Baseline series	−	−	−
4	Balsam of Peru 25%	Baseline series	−	−	−
5	Benzocaine 5%	Baseline series	−	−	−
6	Potassium dichromate 0.5%	Baseline series	−	−	−
7	Ethylenediamine 1%	Baseline series	+	+	+
8	Benzophenone 3%	Baseline series	+	+	+
9	Colophony 20%	Baseline series	−	−	−
10	Chloro-iodo-hydroxyquinoline 5%	Baseline series	−	−	−
11	Tetramethylthiuram disulfide 1%	Baseline series	−	−	−
12	Nickel sulfate 5%	Baseline series	++	++	++
13	Irgasan (triclosan) 2%	Baseline series	−	−	−
14	Formaldehyde 1%	Baseline series	−	−	−
15	Emina lip balm	Patient's own	+	+/−	+/−
16	Emina lip exfoliator	Patient's own	−	−	−
17	L'Oreal liquid foundation	Patient's own	+	+	+
18	Pure Paw Paw lip balm	Patient's own	+	+/−	+/−
19	Sea Make Up facial spray	Patient's own	−	−	−
20	Skintific facial serum (salicylic acid)	Patient's own	−	−	−
21	The Originote sunscreen	Patient's own	+/−	+/−	−
22	Lucienne sunscreen	Patient's own	−	−	−
23	Tamanu facial moisturising cream	Patient's own	+	+/−	+/−
24	Vaseline aloe vera	Patient's own	−	−	−
25	Lucienne facial serum	Patient's own	−	−	−
26	Skintific facial exfoliator	Patient's own	−	−	−
27	Skintific facial toner	Patient's own	−	−	−
28	Skintific facial serum (niacinamide)	Patient's own	−	−	−
29	Sea Make Up micellar water	Patient's own	−	−	−
30	Hada Labo facial exfoliator	Patient's own	+	+	+
31	Skintific facial moisturising cream	Patient's own	−	−	−
32	Sea Make Up face powder	Patient's own	−	−	−
33	Skintific facial moisturising gel	Patient's own	−	−	−
34	N-Pure sunscreen	Patient's own	+/−	−	−
35	Ovale micellar water	Patient's own	−	−	+/−

No	Allergen or product	Series	48 h	72 h	96 h
36	Vaseline Repairing Jelly	Patient's own	+/−	+/−	+
37	White soft paraffin (vaselinum album)	Patient's own	+/−	+/−	+
38	Implora lip balm	Patient's own	+	+	+
39	Dear Me Beauty lipstick	Patient's own	+	+	+
40	Ultimate lipstick	Patient's own	+	+	+
41	Purbasari lipstick	Patient's own	+	+	+/−
42	Pinkflash lip balm	Patient's own	+/−	+/−	+
43	The Originote lip balm	Patient's own	+/−	+/−	+
44	Implora lipstick	Patient's own	+/−	+	+
45	Hada Labo facial wash	Patient's own	−	−	−
46	Chocowhite facial wash	Patient's own	−	−	−
47	Pepsodent toothpaste	Patient's own	−	−	−
48	Whitemilky facial wash	Patient's own	−	−	−
49	Sensodyne toothpaste	Patient's own	−	−	−

Notes: *(−) negative; (+/−) doubtful; (+) weak positive with erythema, mild oedema and non-vesicular papules; (++) strong positive with oedema and vesiculation; (+++) extreme positive with bullae. No unit reached (+++), and no reaction was recorded as irritant (IR) or not tested (NT). Transcribed verbatim from the source clinical record; the grades in this table were reconciled against the narrative description of each reading and agree at every one of the 147 cells.*

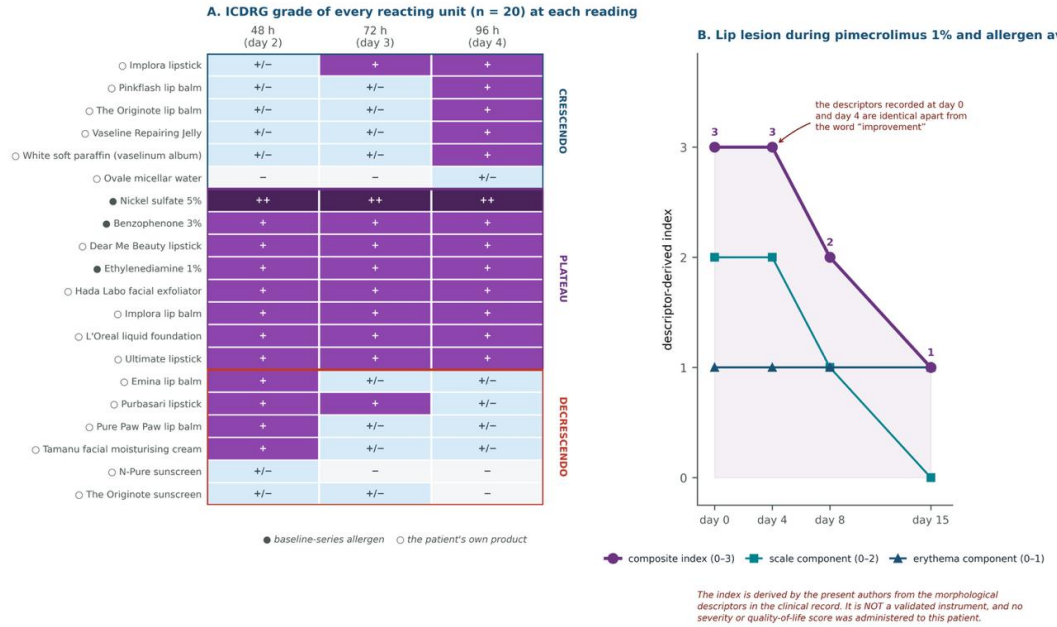

Figure 2. Longitudinal changes in patch-test reaction grade and in the lip lesion. (A) ICDRG grade of each of the 20 reacting units at 48, 72 and 96 hours, grouped by kinetic class; filled circles mark baseline-series allergens and open circles the patient's own products. (B) The lip lesion during pimecrolimus 1% cream and structured allergen avoidance, summarised by the investigator-derived descriptor index. The index is not a validated instrument, and no severity or quality-of-life score was administered to this patient.

3.3. Recovery achieved by each reading strategy

Taking the 17 units that reached a definite positive at one or more reading as the reference, a single 48-hour reading recovered 12 of them — 70.6%, 95% CI 46.9–86.7%; a single 72-hour reading 10, 58.8% (36.0–78.4%); and a single 96-hour reading 13, 76.5% (52.7–90.4%). Of the paired strategies, a 48 + 72-hour schedule recovered 13 of 17, 76.5% (52.7–90.4%), and a 72 + 96-hour schedule 14, 82.4% (59.0–93.8%), while a 48 + 96-hour schedule recovered all 17 units — the same set as the full three-reading schedule, and necessarily so, because no unit in this panel reached a definite positive at 72 hours alone. The recovery achieved by every strategy, with its interval and the units it omits, is given in Table 7 and plotted in panel A of Figure 3. No confidence interval is quoted for those two strategies, because their value is fixed by the definition of the reference set.

The units each truncated strategy would have missed are named individually in panel B of Figure 3, and the identity of those units matters more than their number. Relative to a 48 + 72-hour protocol, the day-4 reading added 4 of 17 definite positives — 23.5%, 95% CI 9.6–47.3%: Vaseline Repairing Jelly, White soft paraffin (vaselinum album), Pinkflash lip balm and The Originote lip balm. All four are leave-on preparations applied to the lips, and two of them are the petrolatum

products the patient had been using as treatment. Conversely, a protocol reading only at 96 hours would have missed 4 units that were definite positive earlier — Emina lip balm, Pure Paw Paw lip balm, Tamanu facial moisturising cream and Purbasari lipstick.

Two different comparisons of definite-positive counts were made, and they answer different questions. Comparing the two reading strategies — a 48 + 72-hour schedule against the full three-reading schedule — the discordance was entirely one-directional: 4 units gained definite-positive status and 0 lost it, giving an exact binomial p = 0.125 on 4 discordant units. The direction is unambiguous, but four one-directional discordant pairs cannot reach conventional significance, because the smallest two-sided p attainable on four such pairs is 0.125. Comparing two single readings with each other gives a weaker picture again: between 72 and 96 hours 4 units gained and 1 lost (p = 0.375), and between 48 and 96 hours 5 gained and 4 lost (p = 1.000). Neither test supports a claim that the day-4 reading significantly increases the number of positives, and no such claim is made. What the panel shows is which units the later reading identified.

Table 7. Recovery achieved by each patch-test reading strategy, against the union of all three readings.

Reading strategy	Units recovered	Recovery (95% CI)	Units missed	Products missed
48 h only	12/17	70.6% (46.9–86.7%)	5	Vaseline Repairing Jelly; White soft paraffin (vaselinum album); Pinkflash lip balm; The Originote lip balm; Implora lipstick
72 h only	10/17	58.8% (36.0–78.4%)	7	Emina lip balm; Pure Paw Paw lip balm; Tamanu facial moisturising cream; Vaseline Repairing Jelly; White soft paraffin (vaselinum album); Pinkflash lip balm; The Originote lip balm
96 h only	13/17	76.5% (52.7–90.4%)	4	Emina lip balm; Pure Paw Paw lip balm; Tamanu facial moisturising cream; Purbasari lipstick
48 + 72 h	13/17	76.5% (52.7–90.4%)	4	Vaseline Repairing Jelly; White soft paraffin (vaselinum album); Pinkflash lip balm; The Originote lip balm
48 + 96 h	17/17	100.0% (definitional)	0	None (reference)
72 + 96 h	14/17	82.4% (59.0–93.8%)	3	Emina lip balm; Pure Paw Paw lip balm; Tamanu facial moisturising cream
48 + 72 + 96 h (full)	17/17	100.0% (definitional)	0	None (reference)

Notes: *Recovery is the number of units reaching ICDRG + or ++ under that strategy as a proportion of the 17 units reaching + or ++ at one or more reading. Wilson score 95% intervals are shown for the truncated strategies only. The reference set is the union of the same three readings and is not an external truth standard: the full three-reading schedule attains 100% by construction, and the 48 + 96-hour schedule attains it because no unit in this panel was definite-positive at 72 hours alone. No interval is quoted for those two rows, because an interval on a value fixed by the definition of the reference set would be meaningless.*

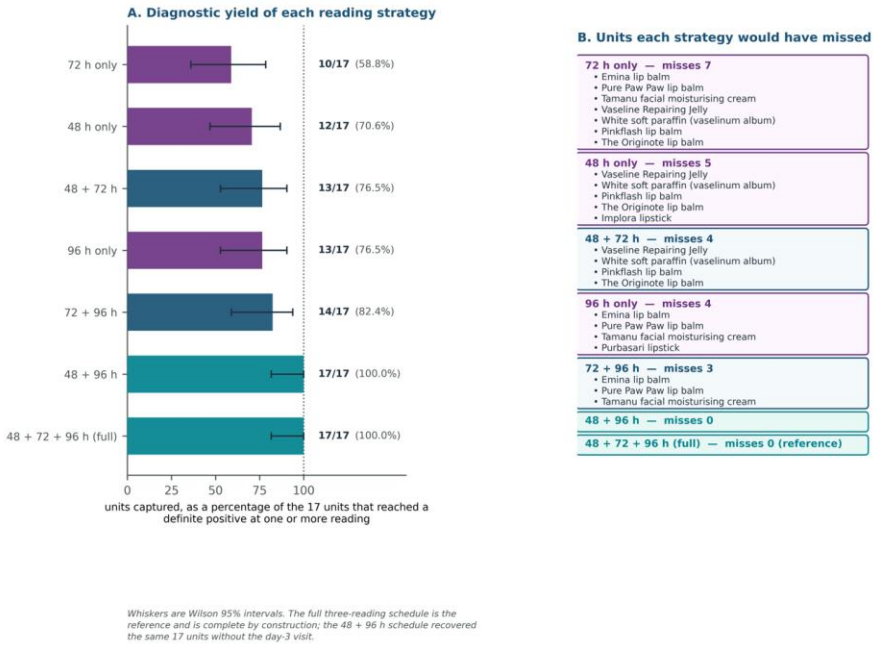

Figure 3. Recovery achieved by each patch-test reading strategy. (A) Proportion of the 17 definite-positive units recovered by each strategy, with Wilson 95% intervals. (B) The units each strategy would have failed to identify. The 48 + 96-hour schedule recovered the same set as the full three-reading schedule, whereas the 48 + 72-hour schedule missed two petrolatum preparations and two lip balms.

3.4. Reaction kinetics

Among the 20 units that produced any reaction, 6 showed crescendo kinetics, 8 plateau kinetics and 6 decrescendo kinetics; no unit was classified as fluctuating as shown in panel A of Figure 2. The crescendo group comprised Ovale micellar water, Vaseline Repairing Jelly, White soft paraffin (vaselinum album), Pinkflash lip balm, The Originote lip balm and Implora lipstick. The decrescendo group comprised Emina lip balm, Pure Paw Paw lip balm, The Originote sunscreen, Tamanu facial moisturising cream, N-Pure sunscreen and Purbasari lipstick. Both baseline-series positives and nickel sulfate were plateau reactions, unchanged across all three readings.

The fate of the doubtful reactions is the practically important finding. Of the 7 units doubtful at 48 hours, 5 (71.4%, 95% CI 35.9–91.8%) had become definite positives by 96 hours, and 2 had faded to negative — the two sunscreens. In the opposite direction, 4 of the 12 units definite positive at 48 hours (33.3%, 95% CI 13.8–60.9%) had fallen back to doubtful by 96 hours: Emina lip balm, Pure Paw Paw lip balm, Tamanu facial moisturising cream and Purbasari lipstick. A doubtful reading at 48 hours was therefore more likely than not to declare itself as a positive by day 4 in this panel, and an early positive was not guaranteed to persist.

3.5. Agreement between reading days

Agreement across the four-level ICDRG scale declined as the interval between readings widened. Linear weighted agreement was highest between 48 and 72 hours (κ = 0.881 (95% CI 0.770–0.972), almost perfect agreement), intermediate between 72 and 96 hours (κ = 0.835 (95% CI 0.708–0.937), almost perfect) and lowest between 48 and 96 hours (κ = 0.725 (95% CI 0.576–0.847), substantial). Quadratically weighted coefficients were uniformly higher, as expected when disagreements are concentrated in adjacent categories: κ = 0.933 (95% CI 0.860–0.984) for 48 versus 72 hours and κ = 0.850 (95% CI 0.744–0.923) for 48 versus 96 hours. Exact grade agreement fell from 89.8% between the first two readings to 75.5% between the first and the last, and agreement on the dichotomised definite-positive classification from 91.8% to 81.6%. All six coefficients, with their bootstrap intervals and their Landis and Koch interpretation, are given in Table 8.

Table 8. Agreement between patch-test reading days across the four-level ICDRG scale (n = 49 units).

Reading pair	Linear weighted κ (95% CI)	Quadratic weighted κ (95% CI)	Exact grade agreement	Agreement on definite positive	Interpretation
48 h vs 72 h	0.881 (0.770–0.972)	0.933 (0.860–0.984)	89.8%	91.8%	Almost perfect
72 h vs 96 h	0.835 (0.708–0.937)	0.909 (0.824–0.967)	85.7%	89.8%	Almost perfect
48 h vs 96 h	0.725 (0.576–0.847)	0.850 (0.744–0.923)	75.5%	81.6%	Substantial

Notes: Bias-corrected percentile bootstrap 95% intervals from 10,000 resamples. Interpretation follows the Landis and Koch bands applied to the linear weighted coefficient. Agreement here is between readings of the same chambers on different days, not between observers.

3.6. Where reactivity concentrated

Reactivity was strongly concentrated in the products the patient applied to her lips. Under the primary classification, 11 of 12 lip-applied products (91.7% (95% CI 64.6–98.5%)) reached a definite positive, against 3 of 23 of her other personal products (13.0% (95% CI 4.5–32.1%)) — a risk difference of 78.6 percentage points (95% CI 45.5 to 89.5), conditional odds ratio 59.0 (95% CI 5.7–3303), Fisher's exact p < 0.001, Cohen's h = 1.82. The pre-specified sensitivity analysis reclassifying the two petrolatum preparations as general emollients attenuated but did not overturn the result (9/10 versus 5/25; odds ratio 31.3, 95% CI 3.2–1653; p < 0.001; h = 1.57).

Vehicle appeared at first to behave in the direction the testing guidance predicts. Leave-on personal products reached a definite positive in 13 of 25 instances (52.0% (95% CI 33.5–70.0%)) against 1 of 10 short-contact products (10.0% (95% CI 1.8–40.4%); odds ratio 9.2, 95% CI 1.0–459; p = 0.028; h = 0.97), and every one of the three facial washes and both toothpastes was negative at all three readings. **That association does not survive stratification by site.** Almost every lip-applied product in the panel is also a leave-on preparation, so the two classifications are largely collinear. Restricted to the patient's non-lip personal products, leave-on preparations reached a definite positive in 2 of 14 instances (14.3%) against 1 of 9 short-contact products (11.1%; odds ratio 1.3, 95% CI 0.06–88; p = 1.000; h = 0.10). The vehicle effect in this panel is therefore not independent evidence: it is a consequence of where the products were applied, and it is not reported as corroboration of the guideline expectation.

The distinction between the patient's own products and the baseline series was not significant: 14 of 35 personal products (40.0%) versus 3 of 14 baseline allergens (21.4%; odds ratio 2.4, 95% CI 0.5–15.8; p = 0.323). What distinguished a reacting unit in this panel was where the product was applied, not whether it came from the standard series. All four comparisons, with their risk differences, odds ratios and effect sizes, are set out in Table 9 and displayed as a forest plot in panel A of Figure 4.

Because the lip-applied classification was made by the investigators, the association was recomputed under 10 alternative but defensible rules, ranging from a strict label-only definition through the exclusion of the petrolatum preparations and of the lip exfoliator to the inclusion of the three facial products as plausibly transferred to the lip; every rule, with its two-by-two table and its result, is given in Table 10. The odds ratio remained above unity under every rule and the association remained significant under every rule

(largest p = 0.019), but its magnitude was highly rule-dependent, ranging from 13.8 to infinity where a classification produced an empty cell. The direction and the presence of the association are robust; the point estimate is not, and should not be quoted without the rule that generated it.

Six hypothesis tests are reported in this manuscript. Under Holm adjustment across all six, exactly 1 survives at the 5% level: the lip-applied comparison (p < 0.001 adjusted). The leave-on comparison does not (0.028 unadjusted, 0.140 adjusted), which reinforces the stratified finding above, and no other comparison approaches significance either before or after adjustment. All six tests are listed with their unadjusted and Holm-adjusted values in Table 11.

Table 9. Concentration of patch-test reactivity by product class.

Comparison	Definite positive, exposed	Definite positive, unexposed	Risk difference (95% CI)	Odds ratio (95% CI)	Cohen's h	p
lip-applied vs other personal product	11/12 (91.7%)	3/23 (13.0%)	78.6 (45.5 to 89.5)	59.0 (5.7–3303)	1.82	< 0.001
S1: petrolatum preparations reclassified as non-lip	9/10 (90.0%)	5/25 (20.0%)	70.0 (34.1 to 83.8)	31.3 (3.2–1653)	1.57	< 0.001
leave-on vs short-contact personal product	13/25 (52.0%)	1/10 (10.0%)	42.0 (6.4 to 61.8)	9.2 (1.0–459)	0.97	0.028
patient's own product vs baseline-series allergen	14/35 (40.0%)	3/14 (21.4%)	18.6 (-11.3 to 40.1)	2.4 (0.5–16)	0.41	0.323

Notes: *Fisher's exact test; conditional maximum-likelihood odds ratio with exact 95% interval; Newcombe hybrid-score interval for the risk difference; risk differences in percentage points. The lip-applied and leave-on classifications were applied by the investigators and are defined in the Methods. The 49 units come from one patient and are not independent; these intervals describe this panel and are not population estimates.*

Table 10. The lip-applied association under ten alternative classification rules.

Classification rule	Definite positive, lip-applied	Definite positive, other	Odds ratio (95% CI)	Cohen's h	p
primary (label names a lip product, or the history records lip application)	11/12 (91.7%)	3/23 (13.0%)	59.0 (5.7–3303)	1.82	< 0.001
S1: the two petrolatum preparations reclassified as general emollients	9/10 (90.0%)	5/25 (20.0%)	31.3 (3.2–1653)	1.57	< 0.001
S2: lip exfoliator excluded (it is a short-contact product)	11/11 (100.0%)	3/24 (12.5%)	inf (9.7–∞)	2.42	< 0.001
S3: lipsticks and lip tints only (balms excluded)	4/4 (100.0%)	10/31 (32.3%)	inf (1.1–∞)	1.93	0.019
S4: balms and emollients only (colour cosmetics excluded)	6/7 (85.7%)	8/28 (28.6%)	13.8 (1.4–723)	1.24	0.010
S5: petrolatum excluded and lip exfoliator excluded	9/9 (100.0%)	5/26 (19.2%)	inf (5.5–∞)	2.23	< 0.001
S6: any product plausibly transferred to the lip, adding the face products	14/15 (93.3%)	0/20 (0.0%)	inf (20.7–∞)	2.62	< 0.001
S7: strict label rule, history disregarded	9/10 (90.0%)	5/25 (20.0%)	31.3 (3.2–1653)	1.57	< 0.001
S8: petrolatum counted as lip-applied, exfoliator as lip-applied, sunscreens added	11/14 (78.6%)	3/21 (14.3%)	19.3 (3.0–186)	1.40	< 0.001
S9: colour cosmetics plus balms, petrolatum and exfoliator excluded	9/9 (100.0%)	5/26 (19.2%)	inf (5.5–∞)	2.23	< 0.001

Notes: *All ten rules were specified by the investigators after the reaction matrix was known, so this is a sensitivity analysis of a post-hoc classification, not a pre-registered robustness check. An infinite upper limit arises where a rule produces an empty cell. The direction and the presence of the association are stable across every rule; the magnitude is not.*

Table 11. Multiplicity: all six hypothesis tests reported in this manuscript.

Test	p (unadjusted)	p (Holm-adjusted)	Survives at 5% after adjustment
lip-applied vs other personal product	< 0.001	< 0.001	Yes
leave-on vs short-contact personal product	0.028	0.140	No
patient's own product vs baseline series	0.323	0.968	No

Test	p (unadjusted)	p (Holm-adjusted)	Survives at 5% after adjustment
definite positives, 48 + 72 h vs full schedule	0.125	0.500	No
definite positives, 72 h vs 96 h	0.375	0.968	No
definite positives, 48 h vs 96 h	1.000	1.000	No

Notes: *Holm step-down adjustment across the six tests. Adjustment is applied to the interpretation and not merely displayed: only the lip-applied comparison is treated as a positive finding in the Discussion and the Conclusion.*

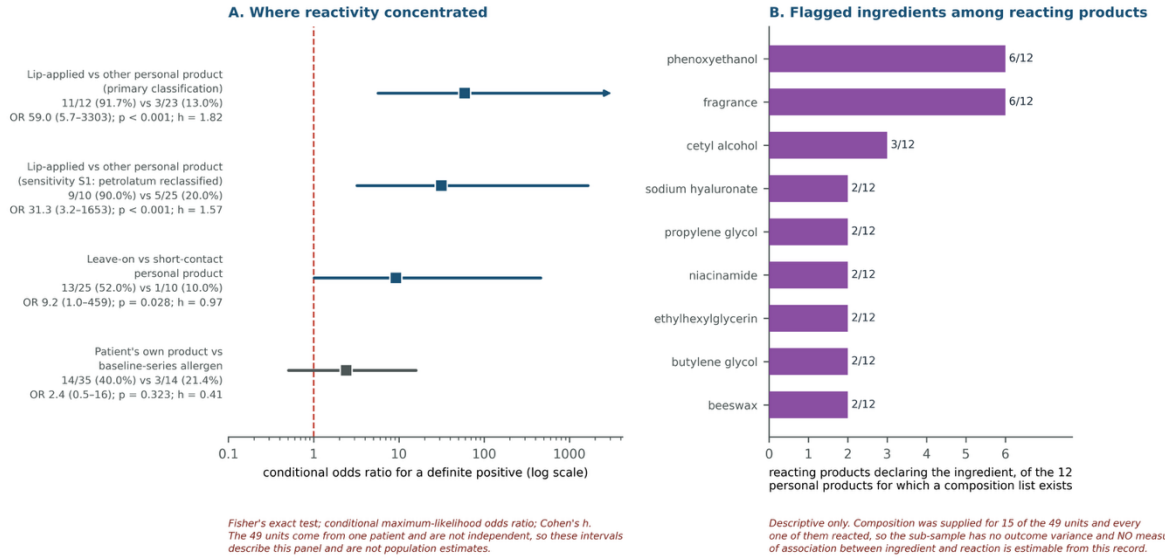

Figure 4. Attribution of patch-test reactivity. (A) Concentration of definite-positive reactions by product class, shown as conditional odds ratios with exact 95% intervals on a logarithmic scale; the arrow indicates an interval extending beyond the plotted range. (B) Frequency with which each flagged ingredient is declared among the reacting products for which the record supplies a composition list. Panel B is descriptive: composition is available for 15 of the 49 units and all 15 reacted, so no measure of association between ingredient and reaction is estimable.

3.7. Declared ingredients and clinical relevance

The record supplies a declared composition for 15 of the 49 units: three baseline allergens tested as single chemicals and 12 of the patient's own products. All 15 produced a non-negative reaction and 13 of the 15 reached a definite positive, while none of the 29 non-reacting personal products has any ingredient data at all. The missingness is therefore informative: composition was transcribed because a unit reacted. There is no denominator against which enrichment of an ingredient among reactors could be judged, and no sensitivity, specificity or odds ratio relating ingredient content to reaction is estimable; ingredient content is reported descriptively only. Among the reacting products, the ingredients flagged in the record as potential allergens most often were fragrance (6 of 12), phenoxyethanol (6 of 12), cetyl alcohol (3 of 12), beeswax (2 of 12), butylene glycol (2 of 12) and ethylhexylglycerin (2 of 12); the full frequency distribution is plotted in panel B of Figure 4 and tabulated in Table 12. One internal inconsistency was found: the record flags cetyl alcohol as the potential allergen in the Tamanu facial moisturising cream, but that product's own declared composition lists cetearyl alcohol and not cetyl alcohol. It is reported rather than harmonised.

Clinical relevance was classifiable as current for 11 of the 17 definite positives: products the patient was using at presentation and applying directly to the vermilion, with a history of aggravation on use and improvement on withdrawal. A further 3 — the L'Oreal liquid foundation, the Hada Labo facial exfoliator and the Tamanu facial moisturising cream — were classified as current (regional) rather than current, because they are facial preparations and the record documents no transfer of them to the lip; the history does record facial swelling in an earlier episode, which makes them clinically relevant to that event but not demonstrably to the cheilitis. The remaining three, all baseline-series allergens, could not be assigned current relevance from this record. Ethylenediamine 1% is declared in none of the 15 products whose composition is recorded. Benzophenone 3% is likewise not declared in any product, although the patient used a sunscreen declaring octocrylene and ethylhexyl methoxycinnamate, so exposure to a chemically related ultraviolet filter is documented even though the tested substance itself is not. Nickel sulfate 5% produced the only strong positive of the entire panel and is declared in no product at all, so the strongest reaction recorded has no documented source. The relevance category assigned to every definite-positive unit, with the basis for each, is given in Table 13.

Table 12. Ingredients flagged in the clinical record as potential allergens, among the 12 personal products for which a declared composition exists.

Ingredient	Reacting products declaring it	Proportion
fragrance	6 of 12	50.0%
phenoxyethanol	6 of 12	50.0%
cetyl alcohol	3 of 12	25.0%
beeswax	2 of 12	16.7%
butylene glycol	2 of 12	16.7%
ethylhexylglycerin	2 of 12	16.7%
niacinamide	2 of 12	16.7%
propylene glycol	2 of 12	16.7%
sodium hyaluronate	2 of 12	16.7%

Notes: *Descriptive frequencies only. Composition was supplied for 15 of the 49 units and every one of them reacted, so this table cannot support any inference that a listed ingredient caused a reaction. Ingredients declared in a single product are omitted from the table and are given in full in the source record.*

Table 13. Clinical relevance of each definite-positive unit, classified post hoc by the investigators.

Unit	Series	Relevance	Basis
Ethylenediamine 1%	Baseline series	Exposed?/Unknown	Ethylenediamine is declared in none of the 15 products whose composition the record supplies; no source of exposure is documented, so current relevance cannot be established from the record.
Benzophenone 3%	Baseline series	Exposed	The patient used a sunscreen declaring octocrylene and ethylhexyl methoxycinnamate; benzophenone itself is not declared in any product, so the reaction is exposure-plausible but not attributable to a declared ingredient.
Nickel sulfate 5%	Baseline series	Unknown	Nickel is declared in none of the products; the strongest reaction of the panel therefore has no documented source and its relevance is undetermined.
Emina lip balm	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.
L'Oreal liquid foundation	Patient's own	Current (regional)	A facial product the patient was using at presentation, applied to the face rather than to the vermilion; transfer to the lip is plausible but is not documented, and the record notes facial swelling in an earlier episode.
Pure Paw Paw lip balm	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.
Tamanu facial moisturising cream	Patient's own	Current (regional)	A facial product the patient was using at presentation, applied to the face rather than to the vermilion; transfer to the lip is plausible but is not documented, and the record notes facial swelling in an earlier episode.
Hada Labo facial exfoliator	Patient's own	Current (regional)	A facial product the patient was using at presentation, applied to the face rather than to the vermilion; transfer to the lip is plausible but is not documented, and the record notes facial swelling in an earlier episode.
Vaseline Repairing Jelly	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.
White soft paraffin (vaselinum album)	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.
Implora lip balm	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.
Dear Me Beauty lipstick	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.
Ultimate lipstick	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.

Unit	Series	Relevance	Basis
Purbasari lipstick	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.
Pinkflash lip balm	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.
The Originote lip balm	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.
Implora lipstick	Patient's own	Current	A product the patient was using at presentation and applying directly to the vermilion, with symptom aggravation on use and improvement on withdrawal.

Notes: The treating team recorded no relevance assessment; this classification was applied by the present authors to the exposure history in the record and is reported as an investigator judgement, not as a recorded finding.

3.8. Treatment and clinical course

Management combined structured education to avoid the products identified as reactive, to avoid licking or wetting the lips with saliva and to maintain lip hydration, with pimecrolimus 1% cream applied twice daily to the vermilion. At day 4 the record described an erythematous patch with overlying thin scale, annotated as improved. At day 8 the erythematous patch persisted with the scale greatly reduced. At day 15 the erythematous patch persisted and the scale had resolved. The investigator-derived descriptor index accordingly fell from 3 at day 0 to 1 at day 15, with the whole of the change accounted for by the scale component; the erythema component was recorded as present at every visit including the last. The trajectory is plotted in panel B of Figure 2, the regimen and its evidence level are set out in Table 14, and the visit-by-visit morphology is given in Table 15.

Two limits of this outcome record should be read alongside it. The descriptors written at day 0 and at day 4 are identical apart from the parenthetical word improvement, so the record cannot distinguish those two visits on morphology alone. And no adverse effect of pimecrolimus — burning, pruritus, irritation, erythema or application-site infection — was recorded at any visit, which is reported here as an absence of recorded events rather than as evidence that none occurred.

Table 14. Treatment protocol, response and evidence level.

Intervention	Regimen	Response recorded	Adverse effects	Evidence level
Structured allergen avoidance	Education to avoid the reactive products, to avoid lip licking and saliva wetting, and to maintain lip hydration	Scale resolved by day 15; erythema persisted	None recorded	Level V (expert consensus and guideline recommendation); the PERDOSKI national guidance and international contact-dermatitis practice both identify allergen elimination as the definitive treatment
Pimecrolimus 1% cream	Applied twice daily to the superior and inferior labial regions from day 0	Descriptor index 3 → 1 between day 0 and day 15	None recorded; no CTCAE grading performed	Level I for the molecule in atopic dermatitis (randomised controlled trial evidence); Level V for this indication, where the evidence is case reports of calcineurin inhibitors in cheilitis
Topical triamcinolone (prior, elsewhere)	Twice daily to the lips before referral	Partial improvement with relapse on withdrawal	None recorded	Level V for this indication; topical corticosteroid is the conventional first line

Notes: Evidence levels follow a conventional I–VII hierarchy in which Level I denotes systematic reviews of randomised trials and Level V denotes descriptive or qualitative evidence and expert consensus. The grading refers to the strength of published evidence for each intervention in this indication and is not a measurement made in this patient. CTCAE: Common Terminology Criteria for Adverse Events.

Table 15. Clinical course of the lip lesion.

Time point	Recorded dermatological status	Erythema	Scale	Descriptor index
Day 0	erythematous patch with overlying thin scale	1	2	3
Day 4	erythematous patch with overlying thin scale (recorded as improved)	1	2	3
Day 8	erythematous patch, scale greatly reduced	1	1	2
Day 15	erythematous patch, scale absent	1	0	1

Notes: *Erythema scored 0 absent, 1 present; scale scored 0 absent, 1 reduced, 2 present. The index is the sum. It was derived by the present authors from the morphological descriptors in the clinical record and is not a validated instrument. The descriptors recorded at day 0 and at day 4 are identical apart from a parenthetical annotation of improvement.*

4. Discussion

This observational analytic case study treated the patch-test chamber rather than the patient as the unit of analysis, and in doing so extracted four findings from a single index case at Dr. Moewardi General Hospital and Sebelas Maret University Hospital in Surakarta. 17 of 49 units reached a definite positive at one or more reading, 82.4% of them the patient's own products. Reaction kinetics divided almost evenly into crescendo, plateau and decrescendo patterns, and the division mapped onto product class in a way that changed the clinical advice given. Agreement between readings fell monotonically as the interval widened, from 0.881 between 48 and 72 hours to 0.725 between 48 and 96 hours. And reactivity concentrated overwhelmingly in lip-applied products, with a risk difference of 78.6 percentage points and an effect size of h = 1.82, robust to the pre-specified reclassification of the petrolatum preparations.

The reading-schedule result is the one with the clearest practical consequence, and it requires care in the telling. A 48 + 72-hour protocol — the schedule most easily delivered when a patient cannot return a fourth time — recovered 13 of 17 definite positives (76.5%), whereas the 48 + 96-hour schedule recommended by the European Society of Contact Dermatitis recovered all of them.[21,22] The four units the earlier schedule would have missed were Vaseline Repairing Jelly, white soft paraffin, Pinkflash lip balm and The Originote lip balm. It would be overstating the evidence to present this as a statistically demonstrated superiority. The discordance between the two strategies is entirely one-directional — four units gained and none lost — but four one-directional pairs give an exact binomial p = 0.125, and 0.125 is the smallest two-sided value four such pairs can produce, so no schedule comparison in a panel this size could have reached significance whatever the result. The weaker single-reading comparisons are further from it still (p = 0.375 for 72 versus 96 hours, p = 1.000 for 48 versus 96 hours). The finding is about identity, not about arithmetic, and it should be read as a description of this panel rather than as demonstrated superiority. Two of the four units missed were the petrolatum-based preparations the patient had herself been applying to her lips as treatment, and the other two were lip balms in daily use. A protocol that stopped at the third day would have returned a result that omitted the products most likely to be perpetuating her disease, and would have left her applying two of them under medical advice.

That reading is consistent with what large series report about late-appearing reactions, and with the physiological reason for them: the effector phase of a type IV response continues to build over 72 to 96 hours, so a reaction graded at 48 hours may be captured mid-ascent (Figure 5).[1,17,22,23] The German S3 consensus and the ESCD guideline both require a reading beyond day 2 for this reason, and both note that the day-4 reading is where crescendo and decrescendo patterns separate.[21-23] What this panel adds is a worked Indonesian example with the missed products named, which is the form in which the argument is likely to persuade a clinic deciding whether the fourth visit is worth its cost.

The kinetic analysis carries the interpretive weight that a single reading cannot. Six units crescendoed, 8 plateaued and six decrescendoed. Crescendo behaviour is the classical signature of an allergic reaction and decrescendo of an irritant one, and in this panel the two groups differ in an informative way.[1,21] Both sunscreens faded from doubtful at 48 hours to negative by 96 hours — a pattern that could also reflect photoallergy, which a conventionally occluded panel cannot detect and which was not investigated here.[51,52] Four units — Emina lip balm, Pure Paw Paw lip balm, the Tamanu facial moisturising cream and Purbasari lipstick — fell from weak positive at 48 hours back to doubtful by 96 hours. Read against the guidance, these six are the reactions a clinician should hesitate to call allergic on the strength of the first reading alone. Conversely, 5 of the 7 units doubtful at 48 hours (71.4%) declared themselves as definite positives by day 4. A doubtful reaction at the first reading was therefore more likely than not to become a positive in this panel, which is an argument against the common practice of dismissing (+/−) results at the first visit.

The declining agreement between readings quantifies the same phenomenon from a different direction. Linear weighted κ between the first two readings was 0.881, in the almost-perfect band, and between the first and the last 0.725, in the substantial band, with the intervals only marginally overlapping. The quadratically weighted values were higher throughout, which tells us that the disagreements were between adjacent grades rather than across the scale — no unit moved from negative to strong positive or the reverse. This matters for interpretation: the day-4 reading did not overturn the day-2 result, it refined it,

moving units by one grade in a direction that depended on their kinetic class. A clinic that reads only once is not making a random error; it is systematically under-calling crescendo reactions and over-calling decrescendo ones.

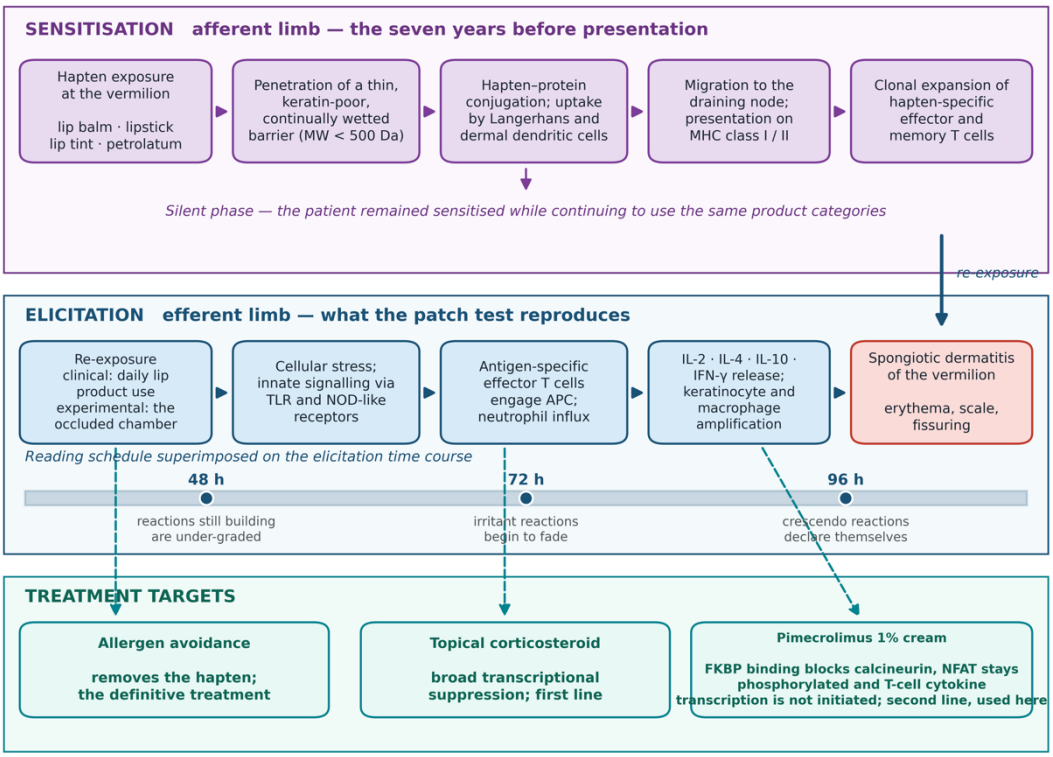

Figure 5. Pathophysiological mechanism of allergic contact cheilitis and the treatment targets used in this patient. The afferent limb produced sensitisation over the seven years before presentation; the efferent limb is what the occluded patch-test chamber reproduces. The reading schedule is superimposed on the elicitation time course to show why crescendo and decrescendo reactions separate only at the later readings. Allergen avoidance removes the hapten, topical corticosteroid suppresses transcription broadly, and pimecrolimus binds FK506-binding protein to block calcineurin, leaving the nuclear factor of activated T cells phosphorylated so that T-cell cytokine transcription is not initiated.

The site-coherence finding is the strongest in the panel and the most directly actionable. 11 of 12 lip-applied products reacted against 3 of 23 of the patient's other products, with a conditional odds ratio of 59.0 and $p < 0.001$. The distinction between the patient's own products and the baseline series, by contrast, was not significant ($p = 0.323$). What separated a reacting unit was where the product was applied, not the provenance of the test substance — a pattern that echoes the region-based diagnostic reasoning recommended in contact dermatitis and the historical attribution of contact cheilitis to lip products in 54% of cases.[2,7,16,19,20,53] The vehicle result appears at first to point the same way — leave-on preparations reacted in 52.0% of instances against 10.0% of short-contact products, and all three facial washes and both toothpastes were negative at every reading — but it does not survive scrutiny and should not be cited as independent corroboration of the guideline expectation that leave-on products are safe to test under occlusion while rinse-off products carry an irritant risk.[22,23,28] Almost every lip-applied product in this panel is also a leave-on preparation, and once the comparison is restricted to the patient's non-lip products the vehicle effect disappears entirely (2/14 versus 1/9; $p = 1.000$). It also fails Holm adjustment across the six tests reported here (0.140). The honest reading is that site and vehicle are collinear in this panel and only site is identified. In this patient the negativity of the toothpastes is itself informative, since toothpaste is the second-commonest cause of contact cheilitis and peppermint-containing dentifrice is a documented culprit.[2,8,9,16]

Two features of this panel require caution before any of it is translated into advice, and both concern the petrolatum result. Petrolatum is the standard vehicle in which patch-test allergens are supplied, and a positive reaction to petrolatum as such is distinctly uncommon; a reaction to a petrolatum-based preparation is more often a reaction to something else in the tube; petrolatum is so standard a vehicle that nickel sulfate itself is conventionally dispersed in it, and comparative work has examined how vehicle and concentration change the reaction obtained.[28,54,55] The record supports that reading here. The white soft paraffin the patient was using declares not only petrolatum but also lanolin alcohol, fragrance and butylated hydroxytoluene, each a documented sensitiser.[17,45,56-58] The Vaseline Repairing Jelly declares petrolatum alone, but a single-ingredient declaration on a consumer label is not an analytical certificate. Neither of these units identifies petrolatum as the hapten, and this manuscript does not claim that it does; what they identify is that two preparations the

patient was applying to her lips as treatment elicited reactions that a day-3 protocol would have missed. The clinical action that follows is to substitute those specific preparations and to test the candidate constituents individually, not to withdraw emollient therapy. Advising a patient with fissured cheilitis to stop using every lip emollient she owns, without providing a tested alternative, would worsen the barrier defect that maintains the disease.

The second caution is the size of the panel. Forty-nine chambers were applied simultaneously, of which 20 reacted. A dense panel with several concurrent strong reactions can produce an excited skin state in which weak reactions appear that would be negative if tested alone, and the classical management is to retest doubtful and weak positives separately after an interval.[21,59] This panel contained only one strong-positive reaction, which makes a marked excited skin state less likely, but it cannot be excluded from the record, no retesting was performed, and the six decrescendo reactions are exactly the pattern that both an irritant response and a regional hyper-reactivity would produce. Every doubtful and decrescendo reaction in this panel should therefore be regarded as provisional until retested individually or confirmed by a repeated open application test.[26-28]

The composition of the baseline arm also bounds what this panel could have found. The 14 allergens tested are a truncated series rather than a full European or North American baseline series, and the omissions are precisely the substances the ingredient declarations implicate: there was no fragrance mix, no lanolin alcohol, no propylene glycol, no paraben mix, no isothiazolinone, and none of the lip-cosmetic-specific allergens such as castor oil, shellac, propolis, beeswax or the azo colourants.[3,21,22,27,32,35,60] A patient sensitised to any of them would have produced positives to her own products, as this one did, while the baseline arm remained silent — which is close to what was observed, and which is the most economical explanation for why 14 of the 17 definite positives were her own preparations. It also means the three baseline positives cannot be read as a complete account of her sensitisations, and that the retest this patient needs is a full baseline series supplemented with the constituents her reacting products declare.

Six hypothesis tests are reported here, and the multiplicity is not incidental to their interpretation. Under Holm adjustment only the lip-applied comparison survives at the 5% level (adjusted $p = 0.00006$); the vehicle comparison falls from 0.028 to 0.140, and nothing else was close before adjustment. The Discussion and the Conclusion therefore treat exactly one association as a positive finding. The lip-applied result is also the only one that survives every alternative classification rule tested, although its magnitude ranges from an odds ratio of 13.8 to infinity across those rules, so the point estimate should never be quoted without the rule that produced it. Readers should treat the kinetic, agreement and recovery analyses as descriptive characterisations of this panel rather than as tested hypotheses.

Three positives could not be assigned current clinical relevance, and the most striking is nickel. Nickel sulfate 5% produced the only strong positive in the entire panel and did so at all three readings, yet nickel is declared in none of the 15 products whose composition the record supplies. Nickel is the commonest positive in essentially every published patch-test series and was positive in 29.8% of the Thai contact-cheilitis cohort, so its presence here is unsurprising; what the record cannot do is connect it to an exposure.[2,32,35,55] Intraoral metal is one documented route to the vermilion, and the history addresses it nowhere.[10] Cosmetic pigments, particularly in coloured lip products, are a recognised route by which trace nickel reaches the vermilion, but no ingredient declaration in this record names it, and asserting that route here would go beyond the evidence. The honest statement is that the strongest reaction in this panel is of undetermined relevance and warrants a targeted exposure history and, if that is unrevealing, a repeated open application test with the suspected products.[21,22,26,27] The same reasoning applies to ethylenediamine, declared in nothing she used, and to benzophenone, where the sunscreen she used declares octocrylene and ethylhexyl methoxycinnamate rather than benzophenone itself.

Ingredient-level attribution is where this record reaches its limit, and it is worth being explicit about why. Composition was supplied for 15 of the 49 units, and every one of those 15 reacted. A sub-sample with no outcome variance cannot yield a sensitivity, a specificity or an odds ratio, however tempting it is to compute one from a table that looks like a two-by-two. Fragrance and phenoxyethanol were each declared in half of the reacting products with a composition list, and lanolin, lanolin alcohol, butylated hydroxytoluene, parabens, propylene glycol and beeswax each appeared in one or two — all substances with documented sensitising potential in the cosmetic literature.[3,17,45,56-58,60,61] But the record supplies no composition for any of the 29 non-reacting personal products, so there is no denominator against which to judge whether these ingredients are enriched among reactors or simply ubiquitous in Indonesian cosmetics. Reporting them as frequencies and refusing the association test is the only defensible treatment, and it identifies exactly what a future study needs: the ingredient lists of the negative products.

The two differentials the treating team recorded were addressed by the same evidence. Exfoliative cheilitis is a chronic, often refractory build-up of keratin on the vermilion of unknown aetiology, sometimes associated with systemic disease, and it presents with dryness and desquamation that overlap closely with this presentation.[1,4,49,62] It was excluded on two grounds: the haematological screen was reported as normal, and an external allergen was identified whose withdrawal coincided with resolution of the scale. Atopic cheilitis, the second differential, arises within the atopic diathesis through barrier dysfunction and immune dysregulation rather than through

sensitisation to a defined hapten, and is usually accompanied by atopic dermatitis elsewhere.[19,53,63] This patient denied personal and family atopy, had no dermatitis at any other site, and produced definite patch-test positives to specific substances, so atopic cheilitis does not account for the findings. It is worth noting what the record leaves open: the history includes symptom aggravation after eating corn and mango, and no food-related testing was performed, so a contributory food-contact component cannot be excluded on this evidence.[3]

Clinically, the case supports the management pathway the PERDOSKI national guidance recommends and the sequence the record followed. Allergen identification came first, avoidance was the definitive intervention, and pharmacological treatment was adjunctive. The choice of pimecrolimus 1% over a further course of topical corticosteroid is defensible on three grounds: the patient had already relapsed twice after triamcinolone; the vermilion tolerates prolonged corticosteroid poorly; and topical calcineurin inhibitors, which block calcineurin-dependent dephosphorylation of the nuclear factor of activated T cells and so prevent transcription of interleukin-2, -4 and -10 and interferon-γ, carry no atrophogenic risk and have cleared cheilitis where corticosteroids failed.[31,47-50,54] The scale resolved by day 15 while the erythema did not, which is the expected sequence when the inflammatory stimulus is withdrawn and barrier recovery lags. It should not, however, be read as a demonstration of drug efficacy: avoidance and pimecrolimus began together, no control condition exists, and the outcome was recorded in morphological descriptors rather than in a scored instrument.

For dermatologists at Dr. Moewardi General Hospital and Sebelas Maret University Hospital, three practice implications follow. First, the patient's own products should be tested, and tested comprehensively: they supplied 82.4% of the definite positives in this panel, and the baseline series alone would have identified three reactive substances, none of which could be traced to anything she was using.[28] Second, the day-4 reading should be protected when a patient's disease involves leave-on products at a permeable site, because that is precisely where the late-declaring reactions were found here. Third, the clinical value of the exercise is the avoidance list it produces, and that list is only as complete as the reading schedule and the product inventory that generated it. The Indonesian context sharpens all three points: patients present with large inventories of locally manufactured lip products at low unit price, whose ingredient declarations are represented in no European or North American baseline series, and for whom no national contact-allergy surveillance data exist to guide which allergens to prioritise.[2]

The study has three strengths. The unit of analysis was defined a priori as the patch-test chamber, which converted a single case into a panel of 49 units with 147 graded observations and made pre-specified hypothesis testing possible without any reconstruction of data. Every reading of every unit was available, so there were no missing observations and the reading-strategy comparison is exact rather than estimated.

Four limitations bound what may be concluded. The 49 units come from one patient and share one immune system, one test field and one occlusion period, so they are not independent observations; the confidence intervals reported here describe the precision with which this panel is summarised and are not estimates of any population quantity, and no result should be transferred to other patients without replication in a multi-patient series. Second, the examination protocol is documented only to the level reported above — chamber diameter, the vehicle and concentration at which the patient's own products were applied, total occlusion time, and the identity and number of readers are not recorded — so the study is not reproducible from this description alone, and a reader cannot exclude the possibility that some doubtful reactions reflected irritancy from an undiluted product. Third, no validated severity or quality-of-life instrument was administered, so treatment response is reported as a descriptor-derived index that the present authors constructed, the descriptors at day 0 and day 4 are indistinguishable apart from an annotation of improvement, and no minimal clinically important difference can be applied; the eight analyses this precludes are enumerated in Table 1. Fourth, clinical relevance was classified after the fact by the present authors rather than by the treating team, no repeated open application test or use test was performed to confirm the doubtful and decrescendo reactions, and follow-up ended at day 15, so neither the durability of the response nor the completeness of the avoidance advice can be judged. Three further constraints belong in the same paragraph. The baseline arm was a 14-allergen truncated series that omits the substances the ingredient declarations most implicate, so the sensitisations recorded here are a lower bound. No skin phototype was recorded, and grading erythema reliably in more deeply pigmented skin is a recognised difficulty for which structured training has had to be developed, so the doubtful and weak-positive grades carry an unquantified reader-dependent uncertainty.[64] The classifications used in the principal analysis were made by the investigators after the reaction matrix was known, so the ten-rule sensitivity analysis is a robustness check on a post-hoc classification and not a pre-registered one. And the recovery analysis evaluates each reading strategy against the union of the same three readings rather than against an external truth standard, so the 100% attained by the full schedule and by the 48 + 96-hour schedule is a property of the definition and carries no information about how many sensitisations the panel detected in total.

5. Conclusion

In this observational analytic case study of allergic contact cheilitis in a 23-year-old Indonesian woman, 17 of 49 patch-test units reached a definite positive at one or more reading and 82.4% of those positives were the

patient's own skin-care and cosmetic products. Reactivity concentrated in lip-applied products (91.7% versus 13.0%; $p < 0.001$; $h = 1.82$) — the only one of six reported tests to survive Holm adjustment, and the only association stable across ten alternative classification rules. A 48 + 72-hour reading schedule would have missed four leave-on lip preparations that the day-4 reading identified, including the two petrolatum-based products the patient was using as treatment; the discordance was entirely one-directional but rests on four pairs and does not reach conventional significance. Weighted agreement between readings fell from $\kappa = 0.881$ to $\kappa = 0.725$ as the interval widened. The lesion cleared of scale by day 15 on pimecrolimus 1% cream and structured avoidance, in keeping with the PERDOSKI national guidance that allergen identification and elimination is the definitive treatment with topical therapy adjunctive. Indonesian dermatology services should test the patient's own product inventory and protect the day-4 reading. Multicentre Indonesian studies with validated severity and quality-of-life instruments, ingredient declarations for non-reacting as well as reacting products, and prospective relevance assessment are needed to establish whether these within-panel patterns hold across patients.

Declarations

Use of Generative Artificial Intelligence. None declared. The authors reviewed and verified all content, and take full responsibility for the integrity and accuracy of the work.

Author Contributions. NK: conceptualisation, clinical assessment, data curation, investigation, writing — original draft. TO: conceptualisation, supervision, clinical assessment, methodology, writing — review and editing, corresponding author. SW: supervision, clinical assessment, methodology, writing — review and editing. WB: investigation, validation, writing — review and editing. DAD: investigation, validation, writing — review and editing. All authors have reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding. The authors received no financial support for the research, authorship or publication of this article. No grant, sponsorship, equipment, materials or other assistance was sought or obtained from any funding agency in the public, commercial or not-for-profit sectors, the patch-test materials were those in routine departmental use, and no organisation had any role in the design of the study, in the analysis or interpretation of the data, in the writing of the report or in the decision to submit it for publication.

Conflict of Interest. The authors declare that they have no competing interests, financial or non-financial, in relation to this work. None of the authors holds any consultancy, employment, shareholding, patent or paid relationship with any manufacturer, distributor or retailer of the products named here. Product brand names appear in this manuscript solely because they are the substances that were tested in this patient. No endorsement or criticism of any manufacturer is intended or implied. A positive patch-test reaction in one sensitised individual is not evidence that a product is unsafe for anyone else, and none of these results should be read as a statement about product quality or regulatory compliance.

Ethical Approval. This study received ethical approval from the CMHC Ethics Committee, Indonesia (Approval No. 2026/325). The study was conducted in accordance with the Declaration of Helsinki. The approval number was supplied by the corresponding author and does not appear in the case documentation from which the clinical data were transcribed; the approval letter is available to the editorial office on request.

Patient Consent. Written informed consent was obtained from the patient for participation and for the publication of her clinical photographs and accompanying clinical data. The case material includes identifiable facial photography, so the signed consent form accompanies this submission; should it not be producible, the facial photograph is to be withdrawn from the article.

Data Availability. The transcribed 49 × 3 patch-test reaction matrix and the complete analysis script that generates every number, table and figure in this manuscript are available from the corresponding author on reasonable request, together with the verification harness described below. Individual clinical data beyond that reported here are not available, in order to protect patient privacy.

Acknowledgements. The authors thank the staff of the Department of Dermatology and Venereology, Dr. Moewardi General Hospital and Sebelas Maret University Hospital, Surakarta, for their assistance, and the patient for consenting to the publication of her clinical details and photographs.

References

1. Utama RFD, Kariosentono H, Dewi AK, et al. The Crescendo Reaction in Patch Testing: A Key Diagnostic Sign in Allergic Contact Cheilitis from Modern Matte Lipsticks. Open Access Indones J Med Rev. 2025;5(6):1637-1649. doi: 10.37275/oaijmr.v5i6.798
2. Kanokrungsee S, Likittanasombat S, Chaweekulrat P, et al. Prevalence and causative allergens of contact cheilitis in Thailand. Contact Dermatitis. 2023;89(5):345-351. doi: 10.1111/cod.14403
3. Aerts O, Pyl J, Mangodt E, et al. Severe contact cheilitis from cera alba and other cosmetic oils, fats and waxes in lip balms. Contact Dermatitis. 2023;88(4):322-323.

   doi: 10.1111/cod.14272
4. Reinfelder M, Seitz A, Paasch U, et al. Frequency of squamous cell carcinomas in vermilionectomy specimens after confirmed actinic cheilitis. J Eur Acad Dermatol Venereol. 2026. Online ahead of print. doi: 10.1111/jdv.70469
5. Kapadia S, Kanji A, Wakelin SH. Contact dermatitis from cera alba in a lip balm–-An overlooked allergen? Contact Dermatitis. 2023;88(4):324-325. doi: 10.1111/cod.14273
6. Ben Salah N, Raiss H, Afli D, et al. Pretty but Painful: Allergic Cheilitis Caused by Phenoxyethanol in a Lip Gloss in a Young Girl. Dermatitis. 2026:17103568251401553.

   doi: 10.1177/17103568251401553
7. Lahouel I, Trimeche K, Smida S, et al. Recurrent cheilitis revealing contact allergy to a lipstick. J Cosmet Dermatol. 2023;23(3):1099-1100. doi: 10.1111/jocd.16080
8. Saracco E, Rashidy N, Borrelli R, et al. Allergic cheilitis due to stannous fluoride-containing toothpaste: First case from Italy and mini-review of previously published cases. Contact Dermatitis. 2024;91(4):371-373. doi: 10.1111/cod.14625
9. George H, Dendooven E, Leysen J, et al. Recurrent allergic contact stomatitis and aphtosis, without cheilitis, due to stannous (tin)-containing toothpastes. Contact Dermatitis. 2023;89(6):509-511. doi: 10.1111/cod.14418
10. Zeng Q, Wang S, Cheng L, et al. Allergic contact cheilitis caused by nickel from stainless steel crowns in primary molars: An easily overlooked intraoral allergen in children? Contact Dermatitis. 2023;89(6):505-507. doi: 10.1111/cod.14413
11. Patel K, Howard M, Tate B. Cheilitis caused by allergic contact dermatitis to cinnamon in chai tea: A case report. Contact Dermatitis. 2022;88(3):239-240. doi: 10.1111/cod.14261

12. Soriano L, Chowdhury M, Cooper S, et al. Current prevalence and relevance of positive patch test reactions to cosmetic and noncosmetic isothiazolinones in the UK. Br J Dermatol. 2021;185(1):223-225. doi: 10.1111/bjd.19898
13. Mangır Ö, Özkaya E. Fragrance and Preservative Contact Allergens in Cosmetic and Household Cleaning Products in Turkey: Variation by Target Population, Product Type and Manufacturing Origin. Contact Dermatitis. 2026. Online ahead of print. doi: 10.1111/cod.70207
14. Sidhu A, Holm J, Bergersen TK, et al. Prevalence and clinical outcome of contact allergy to methylisothiazolinone/methylchloroisothiazolinone in Southern-East Norway: A retrospective case-study. Contact Dermatitis. 2023;88(6):499-501. doi: 10.1111/cod.14308
15. Antelmi A, Trave I, Gallo R, et al. Prevalence of Contact Allergy to Propolis--Testing With Different Propolis Patch Test Materials. Contact Dermatitis. 2025;92(5):349-357. doi: 10.1111/cod.14773
16. Ali Z, Soriano LF. More Than Meets the Eye: Ingredient Prevalence and Patch Test Relevance in UK Micellar Waters. Clin Exp Dermatol. 2026. Online ahead of print. doi: 10.1093/ced/llag373
17. Ljungberg Silic L, Lefevre M, Bergendorff O, et al. Gene profiling reveals a contact allergy signature in most positive Amerchol L-101 patch test reactions. Contact Dermatitis. 2022;87(1):40-52. doi: 10.1111/cod.14077
18. Caf N, Doğan Bİ, Tümtürk M, et al. Epidemiological Analysis of Allergic Contact Dermatitis Cases with Positive Patch-test Results Emerging After the COVID-19 Pandemic. Bagcilar Med Bull. 2025. doi: 10.4274/bmb.galenos.2025.25238
19. Lin SH, Chao YC. Clinical Characteristics and Patch Test Results in 57 Patients with Contact Dermatitis in Southern Taiwan. J Clin Med. 2025;14(7):2291.
doi: 10.3390/jcm14072291
20. Arora K, Mishra Y, Chaudhary PR. Role of Patch Testing With Standard and Cosmetic Series in Patients With Facial Allergic Contact Dermatitis Associated With Everyday Cosmetic Products. Cureus. 2026. Online ahead of print. doi: 10.7759/cureus.114953
21. Bruze M, Svedman C. Clarification and Modification of the International Contact Dermatitis Research Group Classification of Patch Test Reactions on Behalf of the International Contact Dermatitis Research Group. Dermatitis. 2025;36(5):440-446. doi: 10.1089/derm.2024.0365
22. Svendsen SV, Mortz CG. The benefit of late patch test readings in corticosteroid allergy. Contact Dermatitis. 2022;87(5):466-468. doi: 10.1111/cod.14197
23. Ngeywijit T, Tresukosol P, Puangphet P, et al. The outcome of delayed patch test readings on day 7 in Thai population: A prospective analytic study. Indian J Skin Allergy. 2025;4:154-161. doi: 10.25259/ijsa_29_2025
24. Tupker RA, Stapper WGC, Kelder JC. Predictive Factors for Day 7 Positive Patch Test Readings at a Secondary Referral Centre. Skin Health Dis. 2021;2(1):e79. doi: 10.1002/ski2.79
25. Horissian M, Samaan C, Strassner J, et al. Day 2 Patch Testing Does Not Impact Final Diagnosis in Patch Testing Process. Dermatitis. 2024;35(4):407-409.
doi: 10.1089/derm.2023.0173
26. Arora P, Brumley C, Hylwa S. Clinical relevance of doubtful reactions in patch testing: A single-centre retrospective study. Contact Dermatitis. 2024;90(6):607-612.
doi: 10.1111/cod.14526
27. Reeder MJ, Nihal A, Aravamuthan SR, et al. Allergic or Not: Final Interpretation of Doubtful Patch Test Reactions from the North American Contact Dermatitis Group, 2019–2020. Dermatitis. 2024;35(2):138-143.
doi: 10.1089/derm.2023.0285
28. Pruksaeakanan C, Sukakul T, Kanokrungsee S, et al. The Importance of Patch Testing with Patients' Own Cosmetics in Suspected Cosmetic Contact Allergy Patients. Dermatitis. 2025;36(3):e296-e300. doi: 10.1089/derm.2023.0405
29. Navarro-Triviño FJ, Prados-Carmona Á, Ruiz-Villaverde R, et al. Impact of Allergic Contact Dermatitis on Health-Related Quality of Life: A Cross-Sectional Case–Control Study in a Spanish Population. Contact Dermatitis. 2026;94(6):592-602. doi: 10.1111/cod.70116
30. Pollard B, Collis RW, Stahl D, et al. Changes in Product Use and Quality of Life after Patch Testing in Children with Allergic Contact Dermatitis. Dermatitis. 2022;33(5):337-340. doi: 10.1097/der.0000000000000796
31. Sinha S, Sardana K, Panesar S, et al. Comparison of outcomes of azathioprine, leflunomide and allergen avoidance in patients with patch test-positive pigmented contact dermatitis: a randomized comparative study. Clin Exp Dermatol. 2023;49(3):247-254. doi: 10.1093/ced/llad378
32. Uter W, Wilkinson SM, Aerts O, et al. Patch Test Results With the European Baseline Series, 2021/2022--Joint European Results of the ESSCA and the EBS Working Groups of the ESCD, and the GEIDAC. Contact Dermatitis. 2026;95(1):17-32. doi: 10.1111/cod.70134
33. Uter W, Wilkinson SM, Bauer A, et al. Patch Test Results With Additions to the European Baseline Series, 2021/22--Joint European Results of the ESSCA and the EBS Working Groups of the ESCD, and the GEIDAC. Contact Dermatitis. 2026;95(3):276-285. doi: 10.1111/cod.70183
34. Uter W, Wilkinson SM, Aerts O, et al. Patch test results with the European baseline series, 2019/20--Joint European results of the ESSCA and the EBS working groups of the ESCD, and the GEIDAC. Contact Dermatitis. 2022;87(4):343-355. doi: 10.1111/cod.14170
35. Loman L, Uter W, Armario-Hita JC, et al. European Surveillance System on Contact Allergies (ESSCA): Characteristics of patients patch tested and diagnosed with irritant contact dermatitis. Contact Dermatitis. 2021;85(2):186-197. doi: 10.1111/cod.13833
36. Uter W, Wilkinson SM, Aerts O, et al. European patch test results with audit allergens as candidates for inclusion in the European Baseline Series, 2019/20: Joint results of the ESSCA and the EBS working groups of the ESCD, and the GEIDAC. Contact Dermatitis. 2022;86(5):379-389. doi: 10.1111/cod.14059
37. Choi YJ, Byun JY, Choi YW, et al. Analysis of Positive Patch Test Allergens in Allergic Contact Dermatitis Patients with Atopic Dermatitis. Ann Dermatol. 2023;35(4):303. doi: 10.5021/ad.23.001
38. Bilgic A, Bozca BC, Subası GY, et al. Standard Patch Test Results and Clinical Relevance. Indian J Dermatol. 2022;67(3):258-264. doi: 10.4103/ijd.ijd_965_21
39. Kumar TK, Srilakshmi N, Chethana SG, et al. Analysis of Patch Testing for Allergic Contact Dermatitis in Adults Versus Geriatrics: A 7-Year Retrospective Analysis of 342 Positive Patch Test Cases. Indian J Dermatol. 2026. Online ahead of print. doi: 10.4103/ijd.ijd_834_24
40. Bilgic A, Yilmaz O, Uzun S. Dental patch test results and clinical relevance: 10 years of retrospective experience. Turk J Dermatol. 2021;15(4):90. doi: 10.4103/tjd.tjd_76_21
41. Barwari L, Rustemeyer T, Franken SM, et al. Patch test results in a Dutch paediatric population with suspected contact allergy: A retrospective cohort study. Contact Dermatitis. 2022;88(2):120-128. doi: 10.1111/cod.14231
42. Amornruk N, Siranart N, Sittiwattanawong P, et al. The immediate patch test reaction to fragrance in patients with allergic contact dermatitis to fragrance: A prospective study. J Am Acad Dermatol. 2022;87(5):1042-1048. doi: 10.1016/j.jaad.2022.05.053

43. Sukakul T, Bruze M, Mowitz M, et al. Simultaneous patch testing with fragrance markers in the baseline series and the ingredients of fragrance mixes: An update from southern Sweden. Contact Dermatitis. 2022;86(6):514-523. doi: 10.1111/cod.14072
44. Marmgren V, Mowitz M, Zimerson E, et al. Contact allergy to fragrance mix I and its components in individuals with photocontact allergy to ketoprofen. Contact Dermatitis. 2021;85(6):660-670. doi: 10.1111/cod.13958
45. Özkaya E, Kılıç Sayar S. Low Prevalence of Fragrance Contact Allergy among Turkish Population. Dermatitis. 2022;33(5):373-381. doi: 10.1097/der.0000000000000822
46. Ghadiri SJ, Sambi P, Wilkinson SM, et al. Contact sensitisation to shellac: A possible marker of fragrance allergy. Contact Dermatitis. 2023;89(3):205-206. doi: 10.1111/cod.14364
47. Thaçi D, Bräutigam M, Luger T. Corticosteroid-induced skin damage improved with pimecrolimus cream 1% treatment: a 1-year study in adults with mild to moderate atopic dermatitis. J Dermatolog Treat. 2025;36(1):2493931. doi: 10.1080/09546634.2025.2493931
48. Andersen KE. Oleyl Alcohol May Be an Overlooked Allergen in Pimecrolimus Cream. Dermatitis. 2023;34(6):564. doi: 10.1089/derm.2022.0082
49. Santos MCD, Couto MK, Santin GC, et al. Clinical evidence on topical tacrolimus in exfoliative cheilitis: a systematic review. J Oral Diagn. 2026;11. doi: 10.5327/2525-5711.406
50. Kothari R, Pol D, Asnani D, et al. Comparative Study Between Tacrolimus 0.1% Ointment Versus Topical White Soft Petrolatum Jelly in the Treatment of Oral Retinoid-Induced Cheilitis. Indian J Dermatol. 2024;69(2):202. doi: 10.4103/ijd.ijd_67_23
51. Gómez-Martínez S, Brufau-Cochs M, de la Iglesia-Martín J, et al. Long-Term Observations on the European Photopatch Test Baseline Series (EPTBS) in Real Clinical Practice: 11 Years of Results in a Spanish Cohort and Suggestions for an EPTBS Update. Contact Dermatitis. 2024;92(4):277-282. doi: 10.1111/cod.14743
52. Foubert K, Dendooven E, Theunis M, et al. The presence of benzophenone in sunscreens and cosmetics containing the organic UV filter octocrylene: A laboratory study. Contact Dermatitis. 2021;85(1):69-77. doi: 10.1111/cod.13845
53. Pramanik A, Sahoo S, Panda M, et al. Allergic contact dermatitis in children with atopic cheilitis and eyelid dermatitis – A case series. Indian J Skin Allergy. 2025;4:81-84. doi: 10.25259/ijsa_49_2024
54. Napolitano M, Martora F, Antelmi A, et al. Allergic contact dermatitis to petrolatum: An unknown for patch testing. Contact Dermatitis. 2024;90(6):634-635. doi: 10.1111/cod.14537
55. Bach RO, Svendsen MT, Mose KF, et al. A comparison of patch testing with nickel sulfate in TRUE Test and in petrolatum at 2.5% and 5% concentrations. Contact Dermatitis. 2021;86(3):233-234. doi: 10.1111/cod.14013
56. Johnson H. Lanolin: The 2023 American Contact Dermatitis Society Allergen of the Year. Cutis. 2023;112(2):78-81. doi: 10.12788/cutis.0825
57. Sukakul T, Bruze M, Svedman C. Fragrance Contact Allergy – A Review Focusing on Patch Testing. Acta Derm Venereol. 2024;104:adv40332. doi: 10.2340/actadv.v104.40332
58. Sürgün E, Boyvat A. Increased rates of contact allergy to selected preservatives in patients with allergic contact dermatitis in Turkey. Contact Dermatitis. 2023;90(2):110-115. doi: 10.1111/cod.14435
59. Beyzaee AM, Maibach HI. Excited Skin Syndrome (Angry Back), What Do We Know About It? A Review of the Literature. J Cosmet Dermatol. 2026;25(2):e70568. doi: 10.1111/jocd.70568
60. Kocabas G, Ipenburg NA, de Groot A, et al. Results of patch testing propolis in the European baseline series: A 4-year retrospective study. Contact Dermatitis. 2024;91(5):375-378. doi: 10.1111/cod.14678
61. Kanokrungsee S, Dendooven E, Aerts O. Contact Allergy and Allergic Contact Dermatitis From Propylene Glycol and Related Glycols: Cosmetic Skin Sensitisers After All? Contact Dermatitis. 2026. Online ahead of print. doi: 10.1111/cod.70211
62. Ulusoy E, Abacı A. Exfoliative Cheilitis in Childhood: A Successful Treatment with Tacrolimus. J Pediatr Acad. 2023;4(1):39-41. doi: 10.4274/jpea.2023.193
63. Trimeche K, Lahouel I, Belhadjali H, et al. Contact allergy in atopic dermatitis: A prospective study on prevalence, incriminated allergens and clinical insights. Contact Dermatitis. 2023;90(5):514-519. doi: 10.1111/cod.14494
64. Likki S, Black TA, Katta R, et al. Improving Patch Test Interpretation in Skin of Color: Development and Evaluation of an Asynchronous Educational Module. Dermatitis. 2026:17103568261435300. doi: 10.1177/17103568261435300