

辨证结算学

证的判定与复核中九类被吸收的量

胡 敏 著

德麦国际出版社

版权页

书名 辨证结算学

副题 证的判定与复核中九类被吸收的量

著者 胡敏

出版 德麦国际出版社 (Demai Press)

ISBN 978-1-970820-84-3

本书是作者第二部专著，与第一部《治未病立账学》(ISBN 978-1-970820-91-1) 同属一个系列、共用一套论证工序，两书的分界写在本书导论第七节。

版权所有。未经著作权人书面许可，不得以任何方式复制、转载或用于商业用途。学术引用请注明出处，并注明所引内容属于本书主张的哪一层级——本书第十一章与第十三章对此有明确的分层说明，跨层级引用是本书最担心的误用形式。

本书正文约二十四万三千汉字，凡例见导论第十节。

使用声明

本书不是临床工具。

本书提出的九个读数与二十七问，用于诊断记录制度，不用于诊断病人，不用于评价医生，不用于任何个体层面的处置或人事安排。以下五条是使用本书的前提条件，不是建议；完整表述与理由见第十四章第四节。

一、九个读数指的是判定的位置，不指人。 报告不得按医生分组、按个人排名，不得进入任何形式的个人绩效考核。

二、不得追溯问责历史判定。 在一个没有相应输出格式的体系里，当时的填写是唯一可能的输出，责任在体系不在填写者。

三、编码者一致性检验完成之前，不得据本书作任何不利于机构或个人的安排。 一致性系数低于 0.60 不可用，0.60 至 0.75 仅供探索，高于 0.75 方可用于群体层面的比较。本书自己尚未完成这一步，因此按本条，本书当前不支持任何不利安排。

四、弃判权不得与绩效挂钩，无论正向还是负向。 正向奖励导致推诿，负向惩罚导致强行判定。同理，配格率不得设为考核指标，无论正向负向。

五、未采纳意见的记录，不得对被记录者以外的第三方开放用于评价。 它记录的是一位在场者的专业判断，不是他的错误记录。可用于回查，可去标识后用于研究，经本人同意可用于教学；不可用于评价这个人。

本书全部结论均为理论建构与可证伪设计，尚未完成任何一项实地验证。三处欠账写在导论第九节、第十一章第八节与结语第三节，未作隐藏。

作者

胡敏，SDE 学员，人工智能应用实践者。题词：深度思考，自由生活，创造意义。

早年从事技术研发与管理工 作，后转向发生学、人工智能与新国学的交叉研究，现跟随王德生博士研习，并将所学应用于人工智能、教育、健康、婚姻、心理与社会观察等领域。在 SDE Universes 平台设有个人专栏，另辟风险医学与照料生理学、代理接管与内源调节、中医哲思、存在与流变四个独立频道。

他的写作有一道相当稳定的工序：先撤销一个未经审查的先验，再为被遮蔽的机制命名，然后与四到六个敌意最近的邻居逐一勘界，最后给出能排除竞争解释的证伪设计；相当数量的篇目专设一节写明“本文禁止什么”。本书九章是同一道工序，前面多加了一步——**让三个互不通约的学科传统先互相顶撞，再从其中一家自己的材料里找出推翻三家共同假定的那一件。**

他反复回到那个问题是：**当各项指标看起来正常，一个人、一段关系、一份意义，为什么仍在被悄然掏空？**第一部专著《治未病立账学》从“缺栏”回答它；本书从“吞并”回答它——某样东西不是从来没有进过账，而是进过账，又被这一次判定的结论吃掉了。

任职机构、学位与联系方式未列，由作者补。

目 次

版权页.....	3
使用声明.....	4
作者.....	5
前言.....	7
导论.....	12
第一编 判定	23
第一章 让层.....	24
第二章 吞差.....	58
第三章 配格.....	90
第二编 复核	123
第四章 并身.....	124
第五章 摊差.....	157
第六章 填判.....	190
第三编 还原	222
第七章 随程.....	223
第八章 消参.....	254
第九章 同据.....	284
第四编 结算	314
第十章 同一组学科被问了三遍.....	315
第十一章 九个读数放在同一张表上.....	321
第十二章 九句判据合成一份问卷.....	328
第十三章 九栏之间是相乘还是相加.....	334
第十四章 结算的反噬与伦理总则.....	339
结语.....	344
参考文献与文献线索.....	347
附录.....	360
后记.....	366

前言

这本书的起点，是一件我看不懂的事。

几年前，家里一位长辈久病，先后请过三位有经验的中医。三位看的是同一个人、同一段时间、同一批症状——畏寒、肢冷、腰膝酸软、夜尿频、舌淡胖有齿痕、脉沉细。三位辨出来的证不一样，开出来的方也不一样。更让我不解的是：三张方，各有各的效。

我当时的第一反应和大多数人一样：总有两位是错的，只是不知道是哪两位。这个反应看上去天经地义，其实里面压着一个我从来没有检查过的前提——**辨证是一次描述，而描述只能有一个正确答案。**

后来我把三次就诊的记录并排放在一起，才看见另一件事：三位医生认定成立的病机层几乎是同一批，他们的分歧不在“认定哪几层成立”，而在“这一次动哪一层”。三张方各自有效，因为每一张都动对了一层；三个证名不同，因为证名记的是被动的那一层，不是被认定成立的全部。

那么，被认定成立而这一次没有动的那些层，记在哪里？

不记在哪里。病历上什么都没有。

这件事在我心里放了很久。后来我开始留意别人的病历，又撞见两件同样看不懂的。

一件是一份极其整齐的记录：一位慢性病患者每月复诊，连续二十四次，证型一次未改。这份病历在任何一次质控里都会被评优秀——连续、规范、判定稳定。可是把第一次和第二十四次的四诊记录抽出来、遮住日期、交给一位不知情的医生，他很可能判出两个不同的证。中间那二十三次，每一次都是对的——相邻两次之间确实分辨不开；而整体是错的——首末两端确实不是同一个证。没有任何一次可以被追责，因为没有任何一次是错的。

另一件更短。一次会诊，主治判肝郁脾虚，进修医生判湿热中阻，两人各说各的依据，讨论了十分钟。最后写进病历的是四个字。三天后有人调阅这份病历，他看到的就是那四个字。那十分钟去了哪里？它不在病历上，不在处方上，不在任何一份可检索的文件里。如果那位进修医生第二天调走，这份分歧就彻底不存在了——不是被记在别处，是不存在。

这三件事没有一件是失误。三位医生都合格，那份病历可以拿去评优，那次会诊每天都在发生，每一位在场的人都做出了当时能做出的最好的动作。

我不是医生。我早年做技术研发与管理，看惯的东西是记录、流程、字段与接口——一件事做完之后，系统里留下了什么，下一个人接手时能看到什么，出了问题能不能查回去。这个习惯让我在那三张方面前问出了一个大概只有外行才会问的问题：**这三次判定，各自在病历上留下了什么？**

答案是，各留下四个字。

这本书从头到尾没有对任何一次辨证做过对错判断。这不是自谦，是它能够成立的条件。一旦我开始判断谁辨得准，我就必须有临床资格，而我没有；而只要我只问“这一次判定在记录上留下了什么”，我需要的资格就只两样：会翻记录，会数数。九章的九个读数，九句判据，全部建在这两样上——**它们没有一个需要新仪器，没有一个需要新病人，也没有一个需要读者同意我的任何一条判断。**

我的上一本书《治未病立账学》处理的是另一半问题。那本书说：治未病两千年落不了地，未必是因为检测不够灵敏、认识不足或制度不重视，而可能是因为现行的记录制度根本没有为它建立相应的对象与字段——“未”不是时间副词，是记账状态。

写完之后，我发现自己漏掉了一半，而且是更难的那一半。

上一本处理的是**缺栏**：某样东西从来没有进过账。它没有持有者，没有人被指派去做它，也就没有人会为它的缺席负责。这一类的处方相对简单——加一栏。

这本书处理的是**吞并**：某样东西进过账，又被结论自己吃掉了。它有一个曾经在场的持有者。那位提了异议又被写进“肝郁脾虚”四个字里的进修医生在场；那位判不了却必须在诊断栏里填上点什么的医生在场；那位在基础方外加了两味药却没有写理由的医生也在场。他们都看见了，甚至讨论过十分钟。

两者的分界只有一句话：**问这个空白，有没有一个曾经在场的持有者。**指得出的，属于这本书；指不出的，属于上一本。

为什么是现在写它，而不是五年前或五年后？因为“进过账又被吃掉”这件事，只有在我做完上一本、把“账本”这个视角用熟了之后才看得见。上一本教会我的是：不要先问一样东西存不存在，先问它在记录上有没有一个位置。带着这个问题回头看那三张方，我才注意到被让开的那些层；再看那份优秀病历，才注意到二十三次“没有明显变化”累加成的那一段真实变化；再看那十分钟，才注意到它不是没被写好，是没有地方可写。这三样东西在我学会这么问之前，就在我眼前，我看不见。

这条分界有一个对我很不利的后果，我把它写在了导论里，也写在这里：**这本书要求的那九栏，比上一本更容易被填成空的。**加一个“未采纳意见”栏很容易，但只要一次判定仍然必须终止、仍然必须给出一个确定的答案，被吸收的那一份就仍然会在结论形成的瞬间消失，而新加的那一栏会被填上“无异议”。这本书对抗的不是遗漏，是判定这个动作本身的终止性要求。我不确定这一仗打得赢。

九章最初不是为一本书写的。

它们是九次分开进行的实验，每一次都从同一个位置起手：找一个每天都在发生、而没有任何一方觉得需要解释的临床场景——判据是，把它讲给行内人听，对方的第一反应是“这有什么好说的”。然后请三门互不通约的学问各出一条它们最强硬的答法。要求不是三个角度，是**三条互相取消对方的前提**：其中至少有一对，必须构成“一门学问的日课，正是另一门学问诊断为病根的那件事”。三条顶住之后，挖出它们共同站着、却谁也没说出口的那一句。

最后一步是这道工序里最要紧的：**推翻那句共同假定的材料，不从外面找，只从三家中的某一家自己已经掌握、已经发表、只是从来不往这个方向用的成果里找。**

为什么非得如此？因为从外面找反例太容易，也太廉价——一个外行拿着别处的道理来指摘一门学问，最容易被回一句“你不懂”。而当推翻它的那件事，就摆在它自己的教科书里、自己的规范里、自己引以为豪的一段历史里，这句话就回不出来了。九章的九件自毁材料是：化学的保护基；心理物理学里不可分辨关系的不传递；工程学的分组选配与配对件；分析化学在没有标准物质时如何重新生成“算数”的终点；物理学对基本常数做联立平差时那个把矛盾摊进不确定度的扩展因子；细胞学里聚类算法必然给每个细胞分配一个标签；被认证的程序

反过来规定下一批标准物质；自适应控制里参数估计不收敛而闭环跟踪误差仍然趋零；以及免疫学的胸腺阳性选择——选择的判据，有一部分由被选择者所在的系统自己提供。

九次做下来，交出的九样东西彼此不同：让层、吞差、配格、并身、摊差、填判、随程、消参、同据。但它们的形状是一个：**一次判定不是一次读取，它是一次动作；动作有产物，而产物被这个动作自己的结论吸收了。**

我把九篇装成一本书，不是因为篇数凑够了，而是因为其中有一件事，九篇里没有任何一篇说得出来——**这九样，是不是同一样东西？**

如果是，那九章就是九个换了说法的同义反复，这本书应当变薄到只剩一章。如果不是，那必须说清它们凭什么不是，而“看上去不像”不构成理由。

每一章只占一格，谁也回答不了。只有把九章排在一张三行三列的表上——三组毫无交集的学科，横切三个问题——横读一遍，竖读一遍，才看得出：同一组学科被问了三遍，交出三样互不可还原的量，说明它们不是那几门学科的产物；同一个问题被三组学科各答一遍，交出三样内容不同而形状相同的量，说明它们也不是问法逼出来的。第四编做的就是这件事，**这也是这本书之所以是一本书而不是一个文集的全部理由。**

同一处也是我必须交出的地方。这道工序本身，可能就是一台“生产被吸收的量”的机器——把它对准任何题目，都会产出形状相同的东西。这一条我无法自检，因为检验它的条件，就是不知道这本书。第十章第六节写了唯一的出路：另一组研究者，在不知道本书结论的情况下，用另外三组未经预选的学科，把这九个问题重跑一遍。

有几件事这本书不做，我愿意在开头就说死。

它不诊断病人，不评价医生。九个读数的分母全部是判定次数或项数，不是人；把一个以判定为单位的读数按人聚合，得到的数没有可解释的含义，而它会被当成有。它也不测判定的准确性，不测疗效，不告诉任何人应该怎么辨证。它只回答一个问题：这一次判定做了什么，账上留下了什么。九栏俱空，不意味着这次判定错了——它意味着，若这次判定后来被证明错了，没有任何东西可以被回查。

还有一件更不客气的：**我不确定把这些记下来是不是净收益**。一次会诊里的自由讨论，其临床价值有一半来自它是自由说出来的；把每一句异议都变成归档对象，可能得到的是一份完整的记录，和一场再也不会发生的讨论。第十四章把这一条列为这本书应当被整体撤销的第三个条件，而它不依赖我的任何一条论证出错——九笔量可以全部真实，九个读数可以全部算得出来，而记录它们这件事本身仍然是净损失。我没有能力事先排除它。

这本书欠着三笔账。

第一笔：九句判据里最便宜的三条——把一份研究方案里的诊断标准条目表与疗效评价条目表并排放数交集；数一段时期的判定输出里有几次被记为“无从判定”；看一份复核所依据的证据项各自的日期——一个下午可以做完。**我一次也没有做**。

第二笔：二十七问的编码者一致性没有做。两组人各自独立编码一百份既有归档记录即可，不需要任何机构改变流程。在做出来之前，这套问卷只能用于探索，不能用于任何比较。

第三笔：我说得出该改什么，说不出该由谁改。病历书写规范与临床研究报告规范的修订权，不在任何一位读过这本书的临床医生手里，也不在我手里。这个洞的形状很眼熟——它正是第六章指认的那件事：一个本该有归属的位置上，没有人被指派。**我的处方，落在了我自己指认的缺栏里**。

三笔都写在正文里，没有挪进注释。一本书若把自己没做的事藏起来，它对读者的要求就高于它对自己的要求；而这本书通篇在讲的，恰恰是“做过的事要在账上留下痕迹”。

全书的安排是这样的。导论先把三份档案摆出来，把“判定是一次动作”这句话讲清楚，再解释“结算”这个词为什么不是修辞——会计上的结算，要害不在算总数，在于它不允许出现一笔发生过而无处安放的量。正文分四编：第一编问一个「证」是什么（让层、吞差、配格），第二编问它凭什么被复核（并身、摊差、填判），第三编问为什么七十年的还原研究做不成而临床持续有效（随程、消参、同据）。第四编不为前九章辩护，它做的是相反的事——先问这九个读数是不是同一个，写下一条对本书不利的并章规则；再指出九个读数在任何单

一场所都采不齐，因而完整的相关矩阵原理上做不出来；最后把结算本身会引出的三种反噬和五条伦理约束写在明处。九章可以独立读，次序不构成依赖。

最后给几条路。

时间只够两万字的，读导论和结语，那是这本书的两端。想动手验一次的，去做上面那第一笔账里的第一件事——两份条目表，一张纸，不需要新病人，不需要新经费，也不需要同意我的任何一条判断。想吵架的，第十三章、第十四章和附录四写着怎么把这本书推翻，条件都是可裁决的，不需要任何人的许可。

这本书写完之后，我对辨证这件事的敬意反而更重了一些。九章指认的九笔量，没有一笔是因为某个人疏忽而丢的；它们是一套在没有这些字段的条件下运行了两千年、并且真的把病人治好过的体系，为自己的可运行性所付出的代价。指出代价不等于否定它。恰恰相反——一个能在如此稀薄的记录条件下持续有效的体系，值得被认真地清点一次。

感谢王德生博士多年的指教。这本书的方法从他那里学来，其中的错误与莽撞都属于我自己。也感谢那三位医生——他们各自开出了一张有效的方，而我从他们的分歧里，看见了一件他们没有义务替我看见的事。

胡敏

导论 一次判定做了什么，账上留下了什么

一·三份挑不出毛病的档案

第一份。一位七十岁的老人，畏寒、肢冷、腰膝酸软、夜尿频、舌淡胖有齿痕、脉沉细。三位有经验的医生分别辨为脾肾阳虚、肾阳虚水泛、脾阳不振兼水湿内停。三张方各不相同，三张方各自有效。按教科书上那句“证是疾病发展过程中某一阶段病理本质的概括”，这不可能——病理本质只能有一个。于是文献里把这类事处理成两种东西：要么是辨证水平的差异，要么是殊途同归。前一种说法把它当作错误，后一种说法把它当作运气。两种说法都默认：正确答案只有一个，我们只是没能稳定地找到它。

第二份。一位慢性病患者每月复诊一次。第一次辨为脾气虚。第二次，医生看了，与上次比没有明显变化，仍记脾气虚。第三次、第四次，同样。二十四个月，病历上是一条极其整齐的记录：脾气虚，二十四次，一次未改。这份病历在任何一次质控检查里都会被评优秀——连续、规范、判定稳定。现在做一件从来没有人做的事：把第一次的四诊记录与第二十四次的四诊记录抽出来，遮住日期，交给一位不知道它们来自同一个人的医生。他很可能判出两个不同的证。中间的二十三次判定，每一次都是对的——相邻两次之间确实分辨不开；而整体是错的——首末两端确实不是同一个证。没有任何一次可以被追责，因为没有任何一次是错的。

第三份。一位病人，两位医生看。主治判为肝郁脾虚，进修医生判为湿热中阻。两人各说各的依据：一个抓的是脉弦与情志诱因，一个抓的是舌苔黄腻与口苦。讨论十分钟，主治的意见写进病历，开方。三天后有人调阅这份病历，他看到的是四个字：肝郁脾虚。那十分钟去了哪里？它不在病历上，不在处方上，不在任何一份可检索的文件里。如果那位进修医生第二天调走，这份分歧就彻底不存在了——不是被记在别处，是不存在。

三份档案没有一份是失误。第一份是三位合格医生的合格判断，第二份是一份可以拿去评优的病历，第三份是一次正常的、每天都在发生的会诊。三份档案共同指向一件事：**这套体系在每一次判定中都做了一些事，而这些事做完之后，账上只剩下一个结论。**

二·判定不是描述，是一次动作

有一个假定，两千年来没有被单独拿出来检查过，因为它看上去不像一个假定，它看上去就是“辨证”这个词的意思：

辨证是先认清对象是什么，然后再决定拿它怎么办。

先描述，后处置；描述这一步的产物是一个名字，处置这一步的产物是一张方。在这个图景里，“辨”是一次读取，读取不改变被读取的东西，也不留下自己的产物——读完之后世界上多出来的，只有那个名字。

本书要论证的是：这个图景在中医的临床对象上不成立，而且不成立的原因不在中医。**一次判定不是一次读取，它是一次动作。**它在同时被认定成立的若干层里指定了本次动手的那一层（第一章）；它在与上一次的比较中吸收了一份差异（第二章）；它把这个人归进一个格却没有保证格内可互换（第三章）；它所授权的处置在下一次复核之前已经改写了那个要被复核的身体（第四章）；它在形成结论的那一刻消灭了产生结论过程中的分歧（第五章）；它在没有“不适用”这个输出的体系里把不适用填成了适用（第六章）；它在每一轮标准修订里重划了自己要研究的那个对象（第七章）；它把若干本可分开的内部量并进了一个不可再分的组合（第八章）；它用同一批条目既做判定又做验证（第九章）。

九件事的形状是同一个：**动作有产物，而产物被这个动作自己的结论吸收了。**账上留下的是“脾肾阳虚”四个字；被吸收的是——本次不打算动的那几层、二十四个月里逐次吞掉的那些差异、那张方里超出基础方的那几味、七天的药在身体上留下的那部分改变、那十分钟的分歧、那个本该空着的格、被重划掉的旧分类、被并进组合的参数、以及判据与效据重合的那一份信息。

这九样东西没有一样是被藏起来的。它们全都在光天化日之下发生，被在场的人看见，甚至被口头讨论过。它们只是**没有科目**。

一个证名

← 账上留下的全部

图一 这一次判定做了的事，与账上留下的

三·结算

会计上有一个动作叫结算：一段业务结束时，把这段业务发生过的全部收支各归其位，算出余额，再把余额结转到下一期。结算的要害不在算总数，在于它

不允许出现一笔“发生过而无处安放”的量——凡是无处安放的，必须专门开一个科目，哪怕这个科目叫“待处理财产损溢”。

本书借这个词，是因为九章要求的正是这个动作，而不是通常所说的“记录得更全”。记录得更全是在既有科目里填得更细；结算是在发现一笔量没有科目时，先承认它发生过，再决定它该归到哪里，或者专门为它开一栏。

因此本书的书名不是修辞。九章各自指认了一笔在结算表上无处安放的量，各自给出了一个可以在别的行业并排清点的比率，各自给出了一句不含情态词的问话——那句问话的功能，是让人不必去问任何一位医生，只去翻记录就能知道这一笔有没有被结过。

四·母题的三句话

第一句：被吸收的那一份，不是被漏记了。漏记意味着有一个字段而没有填。九章处理的九样东西，问题不在没填，在**没有字段**。一份病历可以填得毫无缺漏而九栏全空，因为表上根本没有这九栏。

第二句：它也不是难测。九个读数没有一个需要新仪器。让层比数的是方里针对的层数与判定认定成立的层数；存异率数的是归档文件里“未采纳意见”栏的内容；同据率把诊断标准的条目表与疗效评价的条目表并排放，数一数交集有几项。第九章那个数，三十年里成千上万项临床研究执行了同一个形式，而它一次也没有被算过——不是因为难算，是因为没有任何一份研究报告规范要求把这两份表分别列出来。

第三句：越正规，它越不可见。这是九章里最不舒服的一条。第二份档案里那份二十四次判定一致的病历，正因为它规范，吞差率才可以是一。德尔菲法删掉共识度最低的条目，而共识度最低的条目恰恰是最难判、最依赖经验、最需要辨别力的那些。信度报告上写着 $\alpha = 0.91$ ，删掉的正是那些老医生真正在用的条目。九章共同指认的不是不规范的后果，是规范化本身在结构上的代价。

五·九栏总表

下表把九章收在一起。第三列是各章给出的读数，第四列是各章那一句不含情态词的问话，全部按原文抄录，未作改写——它们是本书唯一要求读者带走的东西。

章	被这一次判定吸收掉的那一份	读数	判据（原文）
一 让层	认定成立而本次不动的那些层	让层比	这一次的处置针对的项，与这一次判定认定成立的项，数目是否相同？不同的那些记在哪个字段？
二 吞差	一串“判定为相同”里被逐次吸收的差异	吞差率	把第一次记录与最近一次记录直接并排，不看中间，会不会判出同一个证？记录里登记过几次证型改变？
三 配格	归为同类所未保证的那部分——格内位置	配格率	这一张方里，有几味在基础方之外？这几味针对的是这个人的哪一个特征，记在哪个字段？同一时期辨出同一个证的另一位病人，能不能用这一张方？
四 并身	上一版判断的执行产物，并入了被复核的对象	留底率	这一次复核所依据的记录里，有哪几条是在上一次开方之前就已经写下的？
五 摊差	结论形成的那一刻被消灭的分歧	存异率	这次复核归档的文件里，“未采纳意见”有没有一个字段？上一次填的是什么？
六 填判	本该记为“无从判定”而被填成判定的那些次	弃判率	上一次复核的结论表上，“无从判定”有没有一个可填的格？过去一年被填过几次？
七 随程	每一轮研究在结束前把自己的对象重划掉的那一次	随程率	最近 N 次分类变更中，有几次的当日记录里没有出现新的观察项？
八 消参	被并进可辨识组合、因而失去单独身份的那些量	消参率	在这一次判定所依据的全部观测项里，有多少对在过去 N 次记录中始终同时出现、

			同时缺失？
九 同据	判据与效据重合的那一份信息	同据率	这一次判定所用的观测项，与判定这一次有没有效所用的观测项，有几项是同一项？

九句问话有三条共同性质，值得单独指出：一，没有一句带情态词——没有“应当”“可能”“是否合理”，全都是数数或查有无。二，答案必须去查，不能靠回想——问的是记录，不是问人的印象。三，问的是这一次判定，不是这个对象——它们全都不是关于病人的问题，是关于那次判定的问题。

图二 九次吸收发生在一次判定的哪一步

六·三乘三：同一组学科被问了三遍

九章不是九个各自独立的发现凑在一起。它们排在一张表上：

	A 化学·教育测量·文物与艺术品保存	B 物理／计量·心理·工程	C 细胞／免疫·工程·管理
What 一个证是什么	一 让层	二 吞差	三 配格
How 它凭什么被复核	四 并身	五 摊差	六 填判
Why 为什么七十年还原不成而临床有效	七 随程	八 消参	九 同据

这张表是回答“这九样是不是同一样东西”的唯一硬办法。

横读（学科组不变，问题变）：同一组学科被问了三遍，交出三样互不可还原的量。以 A 组为例——化学、教育测量、文物保存这三家，被问“证是什么”时交出让层，被问“凭什么复核”时交出并身，被问“为什么还原不成”时交出随程。三样东西的分母不同、读数不同、可被推翻的方式也不同。这说明它们不是那三门学科的产物：换一个问题，同一组学科给出的不是同一样东西的三种说法。

竖读（问题不变，学科组变）：同一个问题被三组毫无交集的学科各答一遍，交出三样内容不同而形状相同的量。以 How 一行为例——并身在对象侧（被判断者已被改写），摊差在账目侧（分歧被吸收），填判在输出格式侧（没有“不适用”这个格）。三处位置互不重叠，而三处都是“发生过而账上没有科目”。这说明它们也不是问题的措辞逼出来的：换一组学科，同一个问题问出来的不是同样东西。

两向合起来，只剩下一个解释：**被吸收这件事，是“用一次判定去处理一个连续对象”这件事本身的性质，不是学科的性质，也不是问法的性质。**而这一条，九章里没有任何一章说得出来——因为每一章只占一格。

代价必须同时写下来。同一组学科被用了三次，那一组的盲区就被继承了三次；本书第四编须逐组指认这三处继承。另外，工程学在九章里出现了五次（第三、五、六、八、九章），本书自己的叠计比因此至少是五——若那五处所依赖的工程学结论有一处被推翻，五章同时动摇。这笔账记在第十一章。

横读：同一组学科问三题，交出三样内容不同的量
竖读：同一问题问三组无交集的学科 → 交出三样内容不同而形状相同的量

图三 三组学科 × 三个问题的对照设计

七·与《治未病立账学》的分界

本书与作者的上一本书用同一张学科表、同一套论证工序，因此必须在开篇就把两者分开，否则第二本会被读成第一本的续篇。

《治未病立账学》处理的是缺栏：某样东西从来没有进过账。那本书九章指认的量——维持工作、判定线的移动轨迹、被重复计入的余量、界面上的代持、未走完的返程、被学会了的外部参照、非事件的对价、无归属的悬量、须永久供养的残留——共同的形状是：账本按某个单位建，因而这样东西在结构上找不到位置。它的处方是**立账**，是给一栏不存在的东西设一个位置。

本书处理的是吞并：某样东西进过账，又被这一次判定的结论吃掉了。九章指认的量全都真实地发生在判定过程之中，被在场的人看见过、口头讨论过、甚至写进过草稿——那十分钟的分歧确实存在过，那几层确实被认定成立过，那份差异确实被比较过。它们不是从未获得位置，是在**结论形成的那一刻被结论自己吸收了**。它的处方是**结算**，是把已经发生、已被吸收的那一份从结论里剥出来。

两者的判决性区别可以写成一次可执行的检查：**面对一个空白，问它有没有一个曾经在场的持有者**。缺栏没有持有者——没有人做过那份维持工作的记录，因为从来没有人被指派去做；吞并有持有者——那位进修医生在场，他的判断内容不在。凡是能指出“谁在什么时候看见过它”的，属于本书；指不出的，属于上一本。

这条区别有一个直接后果，对读者也对制度：**缺栏可以靠加一个字段解决，吞并不能**。一个“未采纳意见”栏可以立刻加上去，但只要结论仍然要求终止、仍然要求给出一个确定答案，分歧就仍然会在结论形成的瞬间被吸收，而那一栏会被填成空的。第五章与第十四章都要正面处理这件事：**本书的九栏比上一本的九栏更容易被填成空的**，因为它们对抗的不是遗漏，是判定这个动作的终止性要求本身。

判决检查：问这个空白，有没有一个曾经在场的持有者

图四 缺栏与吞并：两本书的分界

八·与最近几家的书级分界

本书的九章各自与本领域近邻逐一划过界。此处只划书一级的、与整本书最容易混同的六家，每一家附一条判定形式。

一、**证候标准化**。它与本书共享全部现象——评价者不一致、同证不同方、跨研究不收敛——但它把这些读作精度问题，处方是把条目切细、阈值定严。判定形式：**在标准版本锁死、条目已细到共识度全部达标的一个体系里，本书的九个读数是否降为零**。本书预测不会，因为九章的量与条目精度无关；第八章更预测规范化在结构上就是激励削减，因而越规范化，结构可辨识性越差。

二、**辨证一致性研究（Kappa 一系）**。它测的是两位医生判得一样不一样。判定形式：**取一份 Kappa 极高的档案，算它的吞差率与存异率**。第二章与第五章共同预测这两个数会同时很高——一致性高恰恰意味着差异被吸收得干净、分歧被处理得彻底。一致性与被吸收量在本书里是同向的，在既有框架里被当作反向的。

三、**循证医学对中医疗效的批评**。它的判断是“以自身量表为结局的疗效不可信”。本书第九章的分离线在于：它不主张病人的好转是假的，而是区分**回路内有效与回路外有效**——前者真实、可重复、对病人有价值，只是不自动传递到

硬终点。判定形式：控制指标的主客观属性之后，同据率是否仍与报告效应量存在剂量—反应关系；且同据率高的体系，其硬终点效应是否接近零而不恶化。若答案是“恶化”，本书与“指标失真”的通行诊断就分不开了。

四、过度诊断与过度治疗。它处理的是判定过多。第六章处理的是判定该缺而不缺——体系里没有“无从判定”这个输出，故所有不适用被填成适用。判定形式：在一个已经系统性降低诊断率的体系里，弃判率是否随之上升。本书预测不会：少判与“记为无从判定”是两件事，前者仍然给出了一个确定输出。

五、机器学习的选择性预测与拒判机制。这是第六章最近的祖宗，必须正面认领：拒判率与弃判率在形式上几乎是同一个量。分离线在于拒判在那里是一个可选项，在这里是一个不存在的输出格式——前者研究“何时应当拒判”，本书问“这套体系允不允许拒判”。判定形式：给出一个已经内置拒判选项的临床判定系统，看它的弃判率是否随之非零。若仍为零，说明缺的不是选项而是别的东西，第六章须改写。

六、加西亚·哈金一系关于“人的类别”的循环效应。这是第七章最近的祖宗。哈金的循环依赖被分类者知道自己被如何分类并据此改变自己；第七章的随程物不依赖这一条——被重划的可以是不会自我理解的东西，驱动方向在识别程序而不在自我认知。判定形式：取一个被分类者完全不知情、且原则上不可能知情的领域（如样品的形态分析），看分类是否仍随程序修订而重划。若不会，第七章须退回哈金的框架内。

九·本书敢让什么判自己输

九章各有证伪条件与写死日期的赌注，汇总在附录。此处只写全书一级的三条，以及本书自己欠的账。

第一条。若九个读数在同一批对象上采齐之后，两两相关系数普遍高于0.8，则九章讲的是同一样东西的九种说法，本书应当并章、变薄。这一条的检验方案在第十一章，并已预注册了五对最可能相关的读数与各自的预测方向。

第二条。若在一个把九栏全部设立起来的机构里，判定质量在两年内没有可测的改善，则本书的实践主张失败——尽管它的诊断主张可以仍然成立。这一条与第一条属于不同层级，不可互相顶替。

第三条。若有另一组研究者，在不知道本书结论的情况下，用另外三组未经预选的学科重跑这九个问题，得到的不是“被吸收的量”而是别的形状，则本书的三乘三设计只是一次事后命名。**本书自己无法执行这一条**，因为执行它的条件就是不知道本书。

本书欠的两笔账，写在这里而不藏在附录。其一，九句判据里最便宜的那几条——抽查十份归档文件看有没有“未采纳意见”栏、把诊断标准条目表与疗效评价条目表并排数交集——一个下午可以做完，**本书一次也没有做**。其二，判据问卷的编码者一致性没有做：两组人各编一百份记录，比任何一条正面预测都便宜，本书同样没有做。这两笔账在结语里会被再提一次。

十·十一约定

一、“本系列”指作者在同一体例下写成的两组作品：《治未病立账学》九章与本书九章。各章中凡出现“本系列”，均指这两组，不指本书一处。

二、《撑平》《阙前距》《叠裕》《代持的病》《欠返》《漏汰》《无事对价》《悬量》《功债》《共载项》是《治未病立账学》的章，各章首次出现时已挂全名，此后只用书名号。

三、《临界回落》《弃录》是本系列之外的同体例单篇，保留原始划界，读者不必查阅。

四、九章可以独立阅读，次序不构成依赖。但第七、八、九三章给出的处方有严格的先后（先加一个回路外的观测，再扩容观测通道，最后记录归类变更），这一条见第九章。

五、九个读数一律是比率，量纲不同而形式相同；但在完成测量不变性检验之前，不作跨领域的数值高低比较。

六、全书不使用任何学派术语。九章原是一个学科对撞的产物，其生成工序写在附录，正文不提。

七、书中所有关于医家、标准与年代的陈述，凡未加注者均为通行说法；可核查的文献线索按章列在书末，引用者请核对原文。

八、九章各自的人机分工声明已合并为全书一处，见后记。

九、各章的节序、读数、判据、削弱、不适用边界、证伪条件、样本纪律与赌注，采用统一体例，措辞刻意一致，便于跨章比对。其中若干句（如「一门学

问的日课，正是另一门学问诊断为病根的那件事」「要挖开它，必须用三边之一自己已经掌握、已经发表、只是从来不住这个方向用的材料」）在多章中原样重复，这是工序的固定表述，不是疏漏；工序全文见附录五。

十、A 组第三门学科在各章中按各自领域惯例出现为艺术哲学、艺术学、归属鉴定与保存修复档案等名目，本书统称「文物与艺术品保存」。正文年份、序数与量词一律用汉字；仅附录三的赌注日期表为表内紧凑体例，用数字。

十一、本书不是临床工具。九个读数用于诊断制度，不用于评价个人；第十四章的五条伦理约束已前置在使用声明页。

第一编 判定

一个「证」是什么

第一章 让层

——一个「证」为什么是一次带处置授权的分层判定：论辨证中“认定成立而本次不动”的那些层为何在三套账上都不设字段，以及体质如何由让层沉淀而成

提 要 「证」是中医全部临床操作的枢纽，也是两千年来定义最不稳的一个词。它与「病」的分界、与「症」的分界、尤其与「体质」的分界，至今没有一条可以在单次诊察中执行的判据。本章认为，困难不在于定义写得不够细，而在于所有现行定义都把辨证当成一次**描述**——先认清对象是什么，再决定拿它怎么办。本章提出一条相反的主张：一个「证」不是一组四诊要素的达标组合，不是一条已经定型的病机路径，也不是这一次医患遭遇共同构成的一个状态，而是一次带处置授权的分层判定——它在同时被认定成立的若干层里指定了本次动手的那一层，并同时把其余各层认定为“成立而本次不动”。这些被让开的层，本章称为**让层**。本章的材料取自三个互不通约的传统：作品同一性的记谱法判据、化学里的选择性与保护基工艺、教育测量里的诊断—补救落差。推翻这一假定的那件材料来自化学自己：一个分子上“有几个反应位点”这个问题，只有相对于一个给定试剂才有答案；而保护基正是“这一次不动它”被物质化的唯一形态。本章给出一句不含情态词的问话：这一次的处置针对的项，与这一次判定认定成立的项，数目是否相同？不同的那些记在哪个字段？由此得到一个三领域量纲不同而比率相同的读数——**让层比**。用它重做四组分界，结论是：症是采集项，病是跨次被反复重新认定的层集，证是本次分子，而**体质是长期居于分母且从不进入分子的那些层的沉淀**。证与体质在单次诊察中原则上不可分，因为它们单次诊察中本来就是同一样东西；分界只能出现在跨次的账上。

一·一个用了两千年而始终没有边界的词

翻开任何一本中医教材，「证」的定义都写得相当明白：证是疾病发展过程中某一阶段病理本质的概括，包括病位、病性、病势和邪正关系。这句话在课堂上没有问题。它在两个地方开始出问题。

第一个地方是病房。一位七十岁的老人，畏寒、肢冷、腰膝酸软、夜尿频、舌淡胖有齿痕、脉沉细。三位有经验的医生分别辨为脾肾阳虚、肾阳虚水泛、脾阳不振兼水湿内停。三张方各不相同，三张方各自有效。按教材的定义，这不可能——“病理本质”只能有一个。于是这类情形在文献里被处理成两种东西：要么是“辨证水平差异”，要么是“殊途同归”。前者是把它当作错误，后者是把它当作运气。两种处理都默认：正确答案只有一个，我们只是没能稳定地找到它。

第二个地方是量表。近三十年，证候标准化做了大量工作：把四诊信息拆成条目，给条目赋权重，设阈值，做聚类，做隐结构分析，做证素提取。这条路走得很扎实，也确实产出了可用的工具。但它一路都在与同一个东西搏斗——**评价者间一致性**。同一段四诊记录交给两位专家，辨出同一个证的比例长期上不去。标准化研究把这个不一致当作**噪声**，认为它来自条目定义模糊、来自采集不规范、来自个人经验差异，因此解决方案是把条目切得更细、把规范定得更严。

本章要问的是：如果它不是噪声呢？

第三个地方最难，也是本章最终要落到的地方：证与体质。阳虚证的条文是畏寒肢冷、神疲乏力、舌淡苔白、脉沉迟无力。阳虚质的条文是平素畏冷、手足不温、喜热饮食、精神不振、舌淡胖嫩、脉沉迟。两段文字放在一起，没有任何一位读者能在不知道来源的情况下把它们分开。教科书给出的分界有三条，每一条都在同一处失灵：

一是**时间尺度**——体质相对稳定，证候随时而变。但一个慢性肾病患者的脾肾阳虚可以十年不变，而一个人的体质在大病一场后可以数月转型。稳定与否是事后才知道的，它不能用来判断眼前这一次看到的是什么。

二是**有病无病**——体质是未病态，证是已病态。这条在逻辑上走不通：中医的“病”很多时候恰恰要靠证来确定，用病来分证与质，是把定义绕回了自身。

三是**可调性**——体质可调而缓，证治而速。这同样是事后判据，而且它偷偷预设了处置方式已经定了。

三条分界法有一个共同的形状：**它们都要求“再看一次”**。变没变、治了动没动、调得快不快，全都需要第二次观察，然后回溯地判断第一次看到的是证还是质。这意味着一件很硬的事：

在单次诊察里，证与体质在原则上不可分。

这不是技艺不精，不是记录不全，不是标准不细。这是一个结构性的结论，而且它到目前为止没有被正面接受过。本章接受它，并且认为一旦接受它，两千年的这道分界题就换了一个位置——它不再是“这一次看到的是哪一类东西”，而是“这一次的账是怎么记的”。

本章的路线是这样的：先请三个互不通约的传统各出一条它们最强的答法（第二节）；找出这三条答法底下那个谁都没说出口的共同假定（第三节）；再从其中一个传统自己的材料里找出推翻它的那一件（第四节）。推翻之后剩下的

东西就是本章的核心主张（第五节）与它的两根轴（第六、七节）。第八节给出一句可以拿去问病历系统而不必问人的问话，第九节把它变成一个可以在三个领域里并排比较的数。第十节回到题目本身，把证与症、与病、与体质的分界重做一遍。其余各节交代兑现、分界、削弱、边界、可以让本章翻车的观测、以及一条写死日期的赌注。

二·三个不肯让步的答法

要问“一样东西凭什么算是同一样东西”，中医不是唯一被这个问题卡住的行当。有三个传统在完全不同的材料上被同一个问题卡了很久，而且各自给出了强硬且互不相容的答案。本章取它们最强的版本，不取它们温和的版本。

二之一·第一条答法：看它此刻呈现出什么，比对一份清单就够了

这一条来自作品同一性的研究。古德曼在《艺术的语言》里做了一个至今没有被真正绕开的区分：有些艺术是**可由记谱定义的**（音乐、文学、建筑图纸），有些是**必须追溯亲笔的**（绘画、雕塑）。前者的同一性判据非常干脆：**只要这一次执行符合乐谱上的每一个音高与时值，它就是那部作品的一个实例**；演奏得再糟，它仍然是那部作品。后者则相反，同一性只能由制作史给出——一幅完美到无法感知区分的摹本，仍然不是原作。

这个立场的锋利之处在于它敢于承认一个几乎所有人都觉得荒谬的推论：一次错了一个音的、无比动人的演奏，不是那部作品；一次音全对的、木讷的演奏，是。古德曼没有回避这个推论，他明确接受它，理由是同一性判据必须是**构成性的**，而构成性属性必须能被一份记谱法从其余一切属性里切出来。切不出来的那些——音色、速度的微差、演奏者的呼吸、那一晚厅里的空气——统统是**非构成性属性**。它们发生了，被听见了，甚至决定了这一次好不好听，但在“这是不是那部作品”的账上，它们是零。

把这一条移到中医上，它对应的正是证候标准化那条路最纯粹的形态：**证是一份清单，四诊要素是清单上的条目，符合阈值即是该证的一个实例**。这条路径不关心这个病人怎么走到今天，也不关心是哪位医生在什么场合看的。它只要求那份清单足够精确。它的主张可以压成一句：**看它此刻呈现出什么样，比对就够了**。

二之二·第二条答法：看它经了哪条路被组织成这样，才够

这一条来自化学。化学在很长一段时间里也以为，看产物就能知道反应。柯廷—哈米特原理终结了这个想法：当反应物的几种构象互变远快于反应本身时，**产物的比例完全不由反应物当下的构象比例决定，而只由通向各产物的过渡态之间的能垒差决定。**也就是说，你手里拿着一瓶产物，做了最精确的分析，你仍然不知道刚才那一瓶反应物里各构象各占多少——那份分布在产物账上不留任何痕迹。

化学由此发展出一整套只为“看路”而存在的方法：同位素标记、交叉实验、动力学同位素效应、哈米特关系、构型保持与翻转的立体化学探针。这些方法有一个共同点——**它们全都不是从产物读出来的，而是必须另做一次实验。**而机理一旦测出来，它就成了预测行为的唯一可靠依据：同一个产物由不同机理生成，在条件改变时会走向完全不同的地方。

化学还知道一件更极端的事：有一类反应发生了，而产物与反应物完全相同。简并重排就是这样——分子内部真实地发生了一次重组，任何产率表、任何色谱图上都是零变化。要看见它，必须先先在分子里放一个同位素标记，事后看那个标记跑到了哪里。**没有标记，它就不是一件事。**

把这一条移到中医上，它对应的是病机分析那一路：同一组症状可以由完全不同的病机路径到达，症状组合本身不足以定证。它的主张可以压成一句：**看它此刻呈现什么没有用，必须知道它经了哪条路。**

二之三·第三条答法：看它此刻纠缠着什么条件，才够

这一条来自教育学。二十世纪八十年代之后，“能力”这个词在教育测量里越来越难写。情境认知那一支的论证很硬：一个人“会不会”某件事，不是他脑袋里的一个量，而是他与一套具体任务、工具、场所、共同体之间的关系；换一套任务环境，读数就变，而且变的不是精度，是所指。测量学这边则从效度的方向逼近同一件事：一份测验的效度不在测验里面，而在**它被怎么用、用了以后发生了什么。**

这一支最不肯让步的地方是：它拒绝承认存在一个“脱离场合的真值”。你不能说“他其实是会的，只是这次没考好”——因为“其实会”这个说法要成立，必须指出它在哪个场合会，而那就又是一个场合。

把这一条移到中医上，它对应的是把辨证看作医患遭遇的产物：证不在病人身上，而在这一次问诊的构造里；换一位医生、换一个诊室、换一种问法，辨出来的就是另一个。它的主张可以压成一句：**脱开这一次的条件场，读不出任何状态。**

二之四·三条为什么不能同时成立

这三条不是三个角度，是三个互相取消前提的判据。

第一条与第二条互相取消：古德曼的判据明确排除制作史（对可记谱的艺术而言，怎么排练出来的与作品同一性无关）；化学的判据说当下的构成读不出它是什么，必须知道路径。若前者对，化学那一整套机理探针是多余的；若后者对，证候清单就永远不可能是判据。

第一条与第三条是更值钱的一种对立：一方每天必须做的那件事，正是另一方**诊断为病根的那件事**。记谱法的全部工作是把构成性属性从执行条件里剥离出来，而情境认知诊断的病，恰恰就是“把能力从情境里剥离出来”这个动作。两边不是意见不同，是一边的地基是另一边的病灶。

第二条与第三条互相取消：化学坚持机理是可以独立测定的客观事实；教育学坚持任何测定本身也在条件场里、也改变被测的东西。同位素标记这件事在化学里是“不扰动地看路”，在教育学的框架里根本不存在这种东西。

三条不能靠“判谁对”来化解，因为它们根本不在同一个语域里作答。要处理它们，只能另找一个把三者关联起来的位置。找这个位置的办法不是折中，是往下挖——挖到三条共同站着的那块地。

三·三条答法共同站着的那块地

三条答法争的是同一件事：“这一次是不是同一样东西”该由什么裁定。一条说由当下呈现的构成性属性裁，一条说由路径裁，一条说由条件场裁。

它们争得很凶，但有一样东西三条答法谁也没有提出来过，因为在它们看来它根本不需要提：

一次判定的结果是对对象的一个描述，它不包含“这一次要动哪一部分”的授权。

这句话念给三边听，三边都会说这还用说吗。

古德曼那一边：作品同一性的判据当然与“要不要修复它、要怎么演它”无关——那是另一个部门的事，属于价值判断，不属于同一性判断。

化学那一边：机理是事实。测出来是什么就是什么，与你接下来要不要优化产率、要不要放大生产没有关系。**机理测定与工艺开发是两个环节**，这一点在化学教育里被反复强调。

教育学那一边更是把它写成了流程图：**先诊断，后补救**。认知诊断评估的标准框架就是这样画的——由作答反应推断属性掌握模式，得到一份诊断报告，然后据此设计补救教学。两个箭头，两个方框，方向不可逆。

三边在“该由谁裁”上势不两立，在“裁完之后才谈动手”这一点上完全一致。而中医的证候研究，从证素辨证到隐结构建模到证候动物模型，全部建立在同一块地上：**先把证辨清楚，再谈论治**。辨证论治这四个字的通常读法，就是这个顺序。

这块地是本章要挖开的那一块。挖它不能从外面搬一个理由来——从外面搬来的理由只是加入了争论，成为第四种立场，而三边可以各自退回自己的阵地。要挖开它，必须用三边之一自己已经掌握、已经发表、只是从来不往这个方向用的材料。

四·推翻它的那件材料，来自化学自己

化学里有一件事，任何一本本科教科书都写，任何一位合成工作者每天都在做，而没有人把它当作一个关于“判定”的命题。

一个分子上“有几个反应位点”，只有相对于一个给定的试剂才有答案。

一个同时含有醛基和酮基的分子，问它“有几个羰基”，答案是二。问它“有几个反应位点”，答案取决于你手里拿的是什么：面对氰基硼氢化钠在特定 pH 下，醛是位点而酮不是；面对某些有机金属试剂，两个都是；面对某个专一性的酶，可能一个都不是。**位点不是分子的属性，是分子与试剂这一对的属性。**

化学不把这当哲学，把它当技艺，叫**化学选择性**。整个精细合成的功夫，一大半花在这上头。

而这件事有一个物质形态，就是**保护基**。

一个多步合成里，分子上常常有好几个官能团在这一步都会反应，而你只想让其中一个反应。做法是：把其余那些**暂时改造成不反应的形式**——羟基做成硅醚，氨基做成氨基甲酸酯，羰基做成缩醛。做完这一步，再把它们脱回来。

现在请注意保护基在做的事情究竟是什么：

它把“这一次不动它”从一个念头变成了一个可以被称量、被表征、被写进实验记录的实体。

被保护的那个羟基，在这一步里“不是”一个羟基——它不参与反应，它在薄层色谱上是另一个点，它的核磁信号移了位。而脱保护之后它“又是”一个羟基。它是什么，取决于这一步的处置计划，而且这个取决关系被物理地实现了出来。

这件材料在化学自己的领域里回答的是另一个问题——它回答的是“怎么做到只让一个位点反应”，是工艺问题。从来没有人把它拿来当作关于“判定是否包含处置授权”的证据。但它就是那件证据，而且它硬到没有讨论余地：**在化学里，“它是什么”这个判定，本来就是相对于“这一次要拿它怎么办”给出的，并且从业者为此付出真实的物质代价——多两步反应，损失一部分收率。**

化学不但知道这件事，还为它吵过一场很大的架。近二十年有一支相当有力的主张，认为保护基是纯粹的账面损耗：它不增加分子的复杂度，只增加步数，理想的合成应当**不用保护基**。这一支主张推动了不少漂亮工作，也确实做成了一些原来被认为必须保护的合成。

这场争论对本章极为重要，而且方向与本章相反——它要消灭的那个东西，恰恰是“这一次不动它”在这个世界上唯一被物质化的记录。本章在第十四节会正面接住它带来的削弱。

现在，这个共同假定被推翻了。三边共同假定“判定是描述，不含处置授权”，而其中一边自己的日常工艺证明：**至少在化学里，判定的内容由这一次的处置意图切出来，并被物质地实现。**

推翻之后，问题就变了。原来的问题是“证该由哪一维裁定”，现在的问题是：

如果一次判定本来就带着“这一次动哪一层”的授权，那么被授权不动的那些层，去了哪里？

五·核心主张：让层

本章的主张是一句三重否定加一个新东西：

一个「证」不是一组四诊要素的达标组合，不是一条已经定型的病机路径，也不是这一次医患遭遇共同构成的一个状态，而是一次带处置授权的分层判定——它在同时被认定成立的若干层里指定了本次动手的那一层，并同时把其余各层认定为“成立而本次不动”。

那些被认定成立、而这一次不动的层，本章称为**让层**。

三个否定各对应一边：第一个否定的是清单判据，第二个否定的是路径判据，第三个否定的是条件场判据。三边都不会同意这条主张——因为三边都会说，你混淆了描述与决策。但三边的证据合起来只能得出它。

现在把“让层”这个词说实。它必须与四样看起来很像的东西分开，而这四样在中医的语汇里都已经有了位置。

让层不是兼证。兼证是被写下来的。病历上写“肝郁脾虚兼血瘀”，血瘀就在那份记录里，它进了字段，它会被下一位医生看到。让层的定义里带着一**条：它不在任何字段里**。一旦被写下来，它就不再是让层——它变成了兼证，而变成兼证之后它通常也就要求被一并处置。

让层不是被排除的候选。医生辨证时会心里过一遍鉴别：这个畏寒，是脾阳虚还是肾阳虚，还是寒湿困脾？最后判定为脾肾阳虚，其余两条被**排除**。被排除的那些是判定为**不成立的**。让层是判定为**成立的**，只是这一次不动它。这两者在账面上都不出现，但它们是两样东西，而且后果完全不同：被排除的那一条如果下次又冒出来，那是当初辨错了；让层如果下次被处置，那是原计划的第二步。

让层不是失治漏诊。漏诊是没看见。让层是看见了、认了、并且做了一个不动它的决定。这个区别不是修辞——它决定了责任归属，也决定了这一层在下次就诊时该不该被重新提出。

让层不是标本缓急里的“本”。《素问》讲“急则治其标，缓则治其本”，讲的是**次序规则**：先做哪个、后做哪个。这条规则说了“先治什么”，但它从来没有要求写下“这一次没治的是什么”。次序是一条医嘱层面的知识，让层是一个**记账层面的空缺**。两者位置不同，本章第十二节会给出一条判决性的对照。

于是让层有一个可以直接执行的识别方式：

在这一次的判定里被认定成立，在这一次的处置里没有被针对，并且在任何结构化字段里找不到。

三条同时满足才是让层。第一条把它与排除分开，第二条把它与本证分开，第三条把它与兼证分开。

五之一·它为什么是一类原来不存在的东西，而不是一个新说法

有一个检验很硬：删掉这个读数之后，它还在不在？

若一样东西删掉读数之后照样存在、只是不好测，那么提出这个读数的人做的是**测量改进**。若删掉读数之后它压根不算一样东西，那么提出读数的人做的是**造出一类东西**。

让层属于后者。理由不是修辞，是三份账可以逐一去查：

病历这一份账：中医电子病历的结构化字段里有“中医诊断（疾病）”“证候”“治法”“方药”。证候一栏通常允许填一个主证，有的允许加兼证。没有任何一个字段的语义是“本次认定成立而不予处置”。医生若想记，只能写进按语或医嘱的自由文本，而自由文本不进任何统计、不被下一次就诊自动带出、不参与任何检索。

证候标准这一份账：各类证候诊断标准的结构是“主症若干、次症若干、舌脉，满足条件即可诊断为某证”。它的输出是一个证名或一组证名。它没有一个位置容纳“同时成立而本次让开”——因为它根本不预设“本次”这回事，它给出的是一个与时机无关的归类。

疗效评价这一份账：证候疗效目前主要靠证候积分——治疗前后把各项症状按轻重赋分，算下降率，据此判定临床痊愈、显效、有效、无效。这份账的分母是全部症状，包括让层上的那些症状。一个医生若这一次只打算动一层，让层上的症状不会改善，它们照样计入分母，于是这个病例的疗效被系统性地低估。而低估的程度与让层的多少直接相关，却没有任何一处记录了让层的多少。

三份账都不设字段。所以在这三份账里，让层不是记录得少，是不算一样东西。删掉本章这个读数，它不会变成“难以测量的东西”，它会回到它现在所在的地方——不存在。

五之二·一个立刻可以做的推论

如果让层这个东西是真的，那么有一件事应当已经在发生，而且应当能被查出来：

同一批病人，接受分层治疗方案的那些，在证候积分法下的疗效读数应当系统地低于接受“一方统治”方案的那些，而这个差与临床结局无关。

这不是一个需要新数据的预测。中医临床研究的既有数据里已经含着它。它只是从来没有被这样切过。

六·两根轴与四个格

一个名字不构成辨别维度。要成为辨别维度，它必须至少有两根轴，而且两根轴的交叉必须给出四个真实存在、可以逐个举例的格。

第一根轴是这一层此刻成不成立。第二根轴是这一次动不动它。

	这一次动它	这一次不动它
此刻成立	① 本证——写在病历“证候”栏里的那一个，方药所针对的那一层	② 让层——认定成立、这一次不予处置、不在任何字段里
此刻不成立	③ 误治——针对了一个并不成立的层，方向错了	④ 正确排除——鉴别时过了一遍，判为不成立，弃置

四格都不是虚设：

第①格是全部现行制度唯一承认的那一格。证候标准描述它，病历记录它，疗效评价考核它，指南规范它。

第③格是不良事件的经典形态，有完整的问责语言（辨证失误、方不对证）。

第④格是鉴别诊断的产物。它同样不留记录，但它的不留记录没有后果——一个被正确排除的东西，本来就不该占位置。

第②格是本章要打的那一格。

六之一·为什么要打的是第②格而不是第③格

选定要打的那一格，有一条判据：**它必须是全部形式审查都能通过的那一格**。一眼就能看出的坏（该做的没做、该记的没记）不需要任何新框架，任何人都看得见。

第②格通过全部形式审查。病历完整、辨证有据、方证相符、治法明确、疗效可评。翻遍现行的病历质控条目，没有一条会因为存在让层而扣分——因为质控条目里根本没有这个对象。

它还满足另外三条签名：

签名一：它在短期指标上表现为改善。 分层处置在短期内往往更有效。一个同时有肝郁、脾虚、血瘀、阴虚的复杂病人，若一次全动，方大而杂，力散而缓；若只动最上一层，方精而专，见效更快。这是临床上被反复验证的经验，也是“急则治其标”的实际依据。**让层多的处方，短期读数常常比让层少的更好看。**

签名二：它的失效不产生任何信号。 一个让层被遗忘时会发生什么？不会有任何报警。病人换了医生，新医生看到的是病历上的一个证名和一张方；他重新辨一次，可能辨出完全不同的层次结构。原来那位医生认定成立而计划下一步处理的那两层，就此消失。**没有任何一处会记录“某层曾被认定成立而从未被处置”**。这不是漏诊——漏诊至少在理论上可被回溯追究，而让层的消失在结构上不可回溯，因为它从未被记为存在。

签名三：它自我加固，越正规化越严重。 这一条最重要。证候标准化程度越高、病历结构化程度越高、诊疗规范越严格，一次判定被允许输出的证名数量就越受限，**让层就越无处安放**。一份要求“每诊必须填写一个主证、至多一个兼证”的电子病历模板，比一份手写的自由格式病历更彻底地消灭让层。**规范化不是解决这个问题，规范化是这个问题的加速器。**

三条签名全部写满，这一格立得住。

六之二·第④格与第②格是本章与另一路工作的分界

第④格（不成立而不动）与第②格（成立而不动）在账面上都是空的，因此极易被混为一谈。区分它们有一条可以执行的方法，而且它需要的只是下一次就诊的记录：

一个被搁置的层，若在下次判定中被重新认定成立并被处置，它是让层；若在下次被判定为从来就不成立，它是被排除的候选。

这条方法把两者的分界从“当时医生心里怎么想的”转移到“后续记录里怎么走的”，因而它可查。

七·一句不含情态词的问话

一个判断若要真正管用，它必须能被压成一句问话，而这句问话必须能拿去问流程文件、日志和考核办法，**不必拿去问人的判断**。凡是出现“应当”“充分”“真正”“恰当”“合理”这类词的问法，都还没有落地。

本章的问话是：

这一次的处置针对的项，与这一次判定认定成立的项，数目是否相同？不同的那些记在哪个字段？

它不含任何情态词。它问的全是可以清点的东西。把它拿到五个不同的地方去问，答案互不相同：

问一份中医门诊病历：答案是——数目通常不同，而不同的那些**没有字段**。若医生想记，只能写进“按语”，而按语不进入任何结构化统计、不被下一次就诊自动带出。

问一份有机合成的实验记录：答案是——数目不同，而不同的那些**记在保护的引入与脱除步骤里**。它们是物质，可以称量，可以在路线图上数出来。化学是本章考察的所有领域里，唯一把这件事物质化了的。

问一份认知诊断评估的报告加它后续的补救教学计划：答案是——数目几乎总是不同。诊断报告会输出完整的属性掌握模式，标出全部未掌握的属性；补救计划只覆盖其中一部分。两份文件并存，差额可以算出来，**没有任何字段记录“这几个为什么不补”**，也没有任何机制保证下一轮会回来补。

问一份文物修复档案：答案是——数目不同，而且**这个差在原则上是可以算出来的**。修复工作有两份文件：现状与病害报告，列出全部识别到的病害；修复处理报告，记录本次实际处理的项。两份相减就是本次未处理项。**这个差有材料、有依据、可清点，但它没有名字，没有人算它，也不进入任何评价**。这是本章找到的最接近让层字段的一份现成账。

问一个软件缺陷跟踪系统：答案是——数目不同，而不同的那些**有字段**。它们被标记为 deferred 或 won't fix，带状态、带负责人、带下一个版本的计划，可以被检索、被统计、被在下一个迭代里重新提出。这是一个反例，而且是

一个极有力的反例：它证明这个字段是可以被造出来的，而且造出来之后系统不会崩溃。

五个答案互不相同。其中至少两个不是从命题推出来的，而是查出来的：文物修复的两份报告之差、软件的 deferred 状态。本章在提出这条命题时并不知道这两处存在。

八·让层比：一个三处量纲不同而比率相同的读数

有了问话，就可以有数。

让层比 = 本次判定认定成立的层数中，本次处置未予针对的层数所占的比例。

它是一个纯粹的比值，没有量纲，因而可以在量纲完全不同的领域之间并排比较。

在中医：分母是这一次辨证中被认定成立的病机层数，分子是其中未被本方针对的层数。取材来源是四诊记录加处方——处方里每一组药对应哪一层，是可以由方剂学知识判读的，而且这个判读的一致性远高于辨证本身的一致性。

在化学：分母是这一步中对所用试剂具有反应性的位点总数，分子是被保护或以其他方式排除在外的位点数。取材来源是合成路线图与实验记录，逐步可数。化学里这个数已经在被间接使用——步数经济性、原子经济性、氧化还原经济性这几个指标，衡量的正是它的代价面。

在教育测量：分母是一次诊断评估中被识别为未掌握的属性数，分子是本轮补救教学未予覆盖的数。取材来源是诊断报告与教学计划，两份文件都是现成的。

在文物与艺术品保存：分母是现状与病害报告中列出的病害项数，分子是本次修复处理报告未涉及的项数。两份报告是行业惯例，差额可以直接算。

四处量纲完全不同——病机层、反应位点、认知属性、病害项——而比率的含义完全相同。

八之一·它必须与既有指标正交，否则白造

一个新读数最容易犯的错误，是它与某个既有的主要指标高度相关。那样的话，它会被立刻读成那个指标的一个代理，白造。检验方法只有一个：举一个“既有指标最好看、而这个数最难看”的真实例子。举不出来就重造。

本章能举出四个。

化学：一个教科书级的全合成，总收率高、立体选择性接近完美、每一步都可重复——而它用了六个保护基，某些步骤的让层比高达 0.8。收率与选择性这两个主指标最好看的那些路线，往往正是让层比最高的那些。这不是巧合，是因果：收率之所以高，正是因为把其余位点都让开了。

中医：一个证候积分下降 82%、被评为“显效”的病例，它的让层比可以是 0——这一次判定只认定了一层，也只动了这一层。而另一个证候积分只下降 31%、被评为“无效”的病例，让层比可以是 0.75——医生这一次本来就只打算动四层里的一层，另外三层上的症状原样不动，全部计入了分母。在证候积分这个主指标上最难看的那个病例，可能是分层判断做得最清楚的那一个。

教育测量：一次诊断评估的模型拟合度极好、分类一致性极高，而后续补救只覆盖了识别出的未掌握属性中的三分之一。拟合度与让层比之间没有任何关系，甚至可以反向——诊断切得越细，识别出的未掌握属性越多，补救能覆盖的比例越低。

文物与艺术品保存：一次修复的可逆性、可识别性、最小干预三项原则执行得最标准的项目，往往正是本次处理项最少的那一个。最小干预这条原则在字面上就是要把让层比推高。

四个例子的方向是一致的：让层比与既有主指标不但正交，而且在若干情形下结构性地反向——既有指标越好看，这个数越大。这一条使它不可能被读成任何现有指标的代理。

八之二·它自带一件伦理

一个比率、一份名单、一个次数，一旦进入制度，几乎必然会被拿去考核人。因此提出它的人有义务把限制写死。本章写死三条：

第一，这个读数指的是位置，不是人。让层比高，指的是“这个判定—处置结构里有几层被让开”，不是“这位医生偷懒”。同一位医生在不同病人身上的让层比会差很多，因为它主要由病情的层次数决定，不由医生的态度决定。任何把让层比直接换算为医师个人评价的做法，都是误用。

第二，不得为了把这个数做好看，而在安全关键的场合制造改动。具体地说：不得为了降低让层比而要求医生一次动完所有层。在化学里这叫无保护基合成，在中医里这叫大方统治，两者在特定条件下都成立，但两者都不能被当作普遍要求。把这个数当作越低越好的指标，是本章明确反对的用法。

第三，已经按这个读数做出不利安排的，必须公开编码规则并给出复核通道。让层比的计算依赖两个判读：哪些层被认定成立、哪一组药针对哪一层。这两个判读都不是唯一的。任何用它做出不利决定的机构，必须公开它用的是哪套判读规则，并允许被评价者提出另一种判读。

三条缺一，这个读数就不该被使用。

九·回到题目：证与症、与病、与体质，重做一遍

前面八节做的是地基。这一节回答题目。

四组分界之所以两千年没有稳定下来，是因为它们一直被当作同一层面上的四类东西在互相区分——好像有一个叫“病”的东西、一个叫「证」的东西、一个叫“症”的东西、一个叫“体质”的东西，要在它们之间划三条线。本章的主张是：它们不在同一层面上，它们是同一个记账结构里的四个不同位置。

一旦接受“辨证是一次带处置授权的分层判定”，四个位置立刻各就各位。

九之一·症：判定的采集项

症是采集项。它是四诊能采到的东西：病人说的、医生看到摸到听到闻到的。口渴、便溏、脉弦、舌红少苔——这些是原始条目。

症与证的分界因此不是“单个与组合”的分界，这是通常的说法，而它不准确。真正的分界是位置的分界：

症是判定的输入项，层是判定的输出项。

一层可以由多个症共同支持，这是常识。但反过来的那一半才是关键，而它通常被忽略：

一个层可以完全没有独立的症。

伏而未发的那一层就是这样。医生根据这个人的病史、体征的组合方式、以及某几项对治疗的反应，认定某一层成立，而这一层此刻并不贡献任何一条可采

集的症状。这样的层在任何“由症状聚类得出证候”的方法里都不可能出现——聚类只能聚出有条目支持的东西。证候标准化研究之所以永远追不上临床辨证，一部分原因就在这里：它的方法在原理上只能捕捉贡献了条目的层。

这也解释了那个被当作噪声的现象。同一段四诊记录交给两位专家，辨出不同的证——如果辨证的输出是一个证名，这是不一致；如果辨证的输出是一个分层加一个授权，那么两位专家可能认定了同一组层，只是把不同的层排到了第一位。这两种情况在现行记录方式下完全无法分辨，因为记录只保留第一位。

这给出一条可查的预测：若在辨证记录中增加“本次认定成立的全部层”这一项，评价者间一致性应当显著上升，而“本次处置针对层”的一致性上升幅度显著小于前者。也就是说，专家们对“有哪几层”的分歧，小于对“先动哪一层”的分歧。这一条与“辨证水平差异”的解释给出相反的预测——按后者，两项一致性应当同步上升。

九之二·病：跨次被反复重新认定的层集

病是跨次对象。它不是某一次判定的产物，它是若干次判定共同指向的那个东西。

用本章的语言：**病是在连续多次判定中被反复重新认定的那个层集合，加上这些层之间相对稳定的支配关系。**

这样一来，“同病异治”与“异病同治”这两句老话就不再是两条需要额外解释的例外，而是这个结构的直接推论：

同病异治——同一个层集合，不同次的判定把不同的层排到了第一位。层集没变，分子变了。消渴病的层集里有阴虚、有燥热、有气虚、有血瘀，上消下消各期，本证不同而层集同。

异病同治——不同的层集合，本次被排到第一位的那一层同构。分子同，分母不同。故补中益气汤可用于脱肛、久泻、崩漏、重症肌无力——四个层集完全不同，而本次被推到前面的那一层是同一个。

请注意这两句推论的形式：它们不需要任何额外的辅助假设。在把证读成“病理本质的概括”的框架里，同病异治与异病同治都需要一整段解释来化解表面矛盾；在本章的框架里，它们是记账结构的两个平凡后果。**能把两条需要解释的例外变成两条平凡推论，是一个框架该做的事。**

九之三·体质：长期居于分母而从不进入分子的那些层的沉淀

这是本章最锋利的一处，也是四组分界里唯一真正的死结。

第一节已经论证过：三条通行的分界法（时间尺度、有病无病、可调性）全部要求“再看一次”，因此在单次诊察里，证与体质原则上不可分。

本章接受这个结论，并且给出理由：

它们在单次诊察里不可分，是因为它们在单次诊察里本来就是同一样东西。

一个层，在这一次被认定成立而未被处置。它是让层。它在下一次仍被认定成立、仍未被处置。它还是让层。它在此后的二十次里每一次都被认定成立、每一次都未被处置。

到某一次为止，人们不再把它叫作让层，开始把它叫作体质。

体质不是另一类东西。**体质是让层的沉淀**——是一个层在足够多次判定中始终居于分母、从未进入分子之后，所获得的一个新名字。

这个说法立刻给出一条可执行的判据，而且它是零情态词的：

取同一位患者连续 N 次就诊的记录，计算某一层被认定成立的次数与被处置的次数之比。这个比值随 N 增大而趋于无界（分母不增）的，是体质；比值有限且分母持续增长的，是证。

这条判据不问“它稳不稳定”，不问“它是不是病态”，不问“它调不调得动”。它只问两个可以清点的数。它把一个两千年的定义问题，换成了一次跨次记录的除法。

它还有一个必然的推论，而这个推论是反直觉的、也是可查的：

体质是可以被制造出来的。一个层被反复让开，它就会变成体质。

这与“体质禀于先天、成于后天、相对稳定”的通行主张形成正面对照。通行主张说体质的稳定性来自它自身的性质；本章说体质的稳定性有一部分来自处置结构——一个层之所以看起来稳定，部分是因为从来没有人动过它。

这条对照可以被判决。取两组慢性病患者，在同一制度环境内（例如同一家医院的同一科室，避免跨机构比较）：一组接受固定主证的治疗方案（每次动同一层），一组接受轮转处置（有计划地在不同次动不同层）。观察若干年后两组的体质分型稳定性。

若轮转组的体质分型稳定性显著低于固定组，而两组的临床结局无系统差异，则本章对，通行主张错。若两组的体质分型稳定性无差异，则通行主张对，本章这一处错。

需要说清楚的是，这条预测有一个前提：轮转必须是有计划的分层处置，不是随意换方。若两组的处置强度、疗程、依从性不可比，这个比较不成立。

九之四·四个位置的一张表

把四组分界收在一起：

位置	在本章的结构里是什么	时间跨度	在账上的形态
症	判定的输入项	单次	四诊记录，条目化，已有完整字段
证	本次判定的分子——被指定动手的那一层	单次	病历“证候”栏，已有字段
让层	本次判定的分母减分子——认定成立而不动的那些层	单次，但指向下一次	无字段
病	跨次被反复重新认定的层集加其支配关系	多次	“中医诊断”栏，字段有而与证的关系未定义
体质	长期居于分母而从不进入分子的层的沉淀	长期	体质量表，独立于病历系统，与证无接口

这张表里，只有一行是空的。而正是这一行的空缺，使其余四行之间的关系无法被写清楚——因为证与体质的差别、证与病的差别，全都是分母与分子的差别，而分母从来没有被记过。

十·在别的地方，这个空缺长什么样

一个跨领域的判断，只有能在别的领域里据以预测一件具体的事，才算兑现；只是“像”，那是类比。本节逐处兑现，共十四处。

一·化学：保护基的数目是让层比的直接读数。预测：一条路线的保护基数目与它的总收率之间，在同类目标分子中应当呈正相关而非负相关——因为保护

正是为了保住收率。这与“保护基是损耗”的直觉相反，而现有的合成路线数据库足以检验它。

二·化学：简并重排是让层的极端形态。 一个层被认定成立、被动了、而账上零变化。它在化学里的对策是同位素标记——**先放一个标记，再看它去了哪里**。对应到中医：若要证明某一层确实被处置了而症状账上无变化，需要一个不在症状账上的指标（生化、影像、或对特定探药的反应）。这解释了为什么“疗效”与“辨证正确”在中医研究里始终对不齐。

三·化学：柯廷—哈米特原理预测症状分布读不出病机分布。 若病机层之间的相互转化快于处置的作用速度，则症状表现的比例完全不反映病机层的实际比例。这给出一条限制：**症状聚类只在“层间转化慢于观察间隔”时才能逼近病机结构**。这一条可以由重复四诊的间隔实验检验——缩短观察间隔，若聚类结构显著改变，则转化不慢。

四·化学：无保护基合成的追求，是把让层比强行推向零。 它在化学里成功的条件是明确的——目标分子的官能团差异足够大，或试剂的选择性足够高。对应的预测：**中医的大方统治（一次动完所有层）应当只在层间药性冲突小的情形下优于分层处置**。这可由复方规模与疗效的关系检验，且预测的是一条有拐点的曲线，不是单调关系。

五·教育测量：诊断报告与补救计划之间的差额，是教育里的让层。 预测：在诊断—补救之间引入一个“本轮不补”的显式字段（记录哪些属性被识别为未掌握而本轮不予处理），下一轮诊断的属性掌握模式预测精度应当上升。理由是这个字段消除了“上一轮到底教了什么”这一项混淆。这个实验成本极低，只需在现有报告模板上加一列。

六·教育测量：Q 矩阵的属性划分，本身就是按可教学单元切的。 这是教育测量公开承认的设计原则，不是它的盲区。它给出的预测是：**属性的粒度会随补救手段的变化而变化**——当一种新的教学技术使某两个属性可以被一并处理时，后续版本的 Q 矩阵会倾向于把它们合并。这一条可以在同一测验工具的历次修订史上直接查。

七·教育：分层教学与差异化教学是“有意的不覆盖”的制度化形态。 它们有明确的教案，但通常不记录“本轮对某组学生的哪些已识别缺口不予处理”。预测：记录了这一项的学校，其学期末的补救覆盖率应当高于未记录的学校，而两者的初始诊断精度无差异。

八·教育：迁移测验是“必须扰动才能分开”的教育版本。要区分“会了”与“这一次碰巧对了”，只能换一个表面不同、结构相同的任务再看一次。这与本章关于证与体质只能跨次分开的结论，形式完全相同：**两样在单次观察中不可分的东西，唯一的分法是施加一次变化并观察回应。**

九·作品同一性：记谱法把非构成性属性判为零，这正是证候标准的结构。预测：任何一套证候诊断标准，在被严格执行时，都会系统性地丢失那些不贡献标准条目的层。这可以由“按标准辨证”与“专家自由辨证”的对照直接检验——两者的差异不应当随机分布，而应当集中在无独立症状支持的层上。

十·文物与艺术品保存：现状病害报告与修复处理报告之差，是本章找到的最接近让层字段的现成账。它有材料、可清点，但没有名字、没有人算它、不进入任何评价。预测：把这个差显式计算并纳入档案的机构，其后续修复中“复发性病害”的记录率应当上升——不是因为病害变多了，而是因为原本被静默遗忘的项被重新提出。

十一·文物与艺术品保存：最小干预原则在字面上就是把让层比推高的规范。这构成对本章的一个有力提醒：**让层比高不等于坏。**在可逆性和可识别性被高度重视的领域，高让层比是正确操作的标志。这一点在第十一节会被展开为对本章的一处削弱。

十二·软件工程：缺陷跟踪系统里的 deferred 状态，证明这个字段可以被造出来。它带状态、带负责人、带目标版本，可检索、可统计、可在下一轮被自动提出。预测：中医电子病历若增加一个结构相同的字段（本次认定成立而不予处置，带下一次自动带出的标记），则复诊时该层被重新提及的比率应当显著上升，而门诊时长的增加应当小于两分钟。

十三·急诊分诊：分诊也是“这一次只动一部分”，而且它有记录。但它排的是同一批对象之间的先后，本章排的是同一个对象内部各层的先后。这个区别给出一条判决性的对照：分诊记录能回答“没被处理的是哪几个病人”，而中医病历不能回答“没被处理的是哪几层”。同一个结构，一处有账一处无账，差别不在难度，在有沒有人要求过。

十四·精神医学的主诊断与共病编码：制度只认一个主诊断。共病虽有编码位置，但报销、统计与疗效评价通常锚定主诊断。预测：在共病编码被完整要求的制度下，主诊断的分布会比在只要求主诊断的制度下更集中——因为医生会把

原本会分散写出的层，压回到一个主诊断里。这一条可以在同一国家的编码规则修订前后直接查。

十四处兑现里，有三处（第五、第十、第十二）给出的是可以立即执行、成本极低的实验；有两处（第四、第十一）反过来限制了本章的主张。

十一·与最近的几种说法的分界

一个新提出的东西，最大的风险不是被反驳，是被吸收——读者读完说“这不就是某某吗”，于是它作为一个独立对象消失。本节逐一处理最容易发生这种吸收的十种说法。每一种都给出一条可分辨的差别，以及一条能分出对错的观察。

兼证与主次证（中医内部）。兼证说到这一步：一次辨证可以输出不止一个证名，主要的写在前面，次要的写在后面。分离线在于**记不记**：兼证是被写下来的，让层的定义里就带着“不在任何字段里”。这不是程度差别，是有无差别，而且它有一个直接后果——兼证一旦写下，通常就要求被一并处置，方里要加针对它的药；让层写下就不再是让层。**若观察到：病历上写了兼证的病例，其处方中针对兼证的药物在多数情形下确实存在，则兼证与让层是两个对象，本章对；若观察到写了兼证而处方中系统性地不针对它，则兼证栏实际上已经在承担让层的功能，本章这一条被吸收。**这是一个只需要抽查处方就能完成的检验。

标本缓急（《素问·标本病传论》）。它说到这一步：先治哪个、后治哪个，有一套依病势缓急而定的次序规则。这是中医里与本章最近的一条，而且它比本章早两千年。分离线在于它是**次序规则，本章是记账结构**——标本缓急给出了先动哪一层的依据，但它从未要求写下“这一次没动的是哪几层”。次序知识存在于医者的判断里，账目空缺存在于记录系统里，两者可以同时为真。**若观察到：按标本缓急施治的病例，在换医之后其“本”的那一层被新医重新提出的比率与随机重辨没有差异，则本章对——次序知识没有被传下去，因为它没有被记下来；若观察到重新提出的比率显著高于随机，则次序知识确实经由某种途径传递了，本章关于“不可回溯”的主张需要削弱。**

构成性属性与非构成性属性（古德曼，《艺术的语言》，1968，艺术哲学）。它说到这一步：同一性判据必须由一份记谱法把构成性属性从其余属性中切出来，切不出的一律不参与同一性。分离线在于**让层不是非构成性的——它是构成性的，只是被延期。**一个被认定成立的病机层，若下一次被处置，它当然构

成这个病人的病理结构；它这一次不进入判定，不是因为它无关，是因为它被排在了后面。同一份病历，古德曼式的判据判“这是该证的一个完整实例，记录无缺”；本章判“这份记录缺一个字段”。两种判读对同一份材料给出相反的评价，因此分离线成立。

认知诊断模型里的失误与猜测参数（Tatsuoka 1983；de la Torre 2009，教育测量）。这是本章最危险的一位方法学占位者。它说到这一步：作答反应与真实掌握状态之间存在偏离，模型用两个参数把它估计掉——掌握了却答错叫失误，没掌握却答对叫猜测。分离线在于让层不是估计误差，是判定的一个组成部分。失误与猜测是被当作噪声吸收的东西，模型的目标是把它们的方差压小；让层是一个有意做出的决定，它的目标不是被压小，是被记下来。若观察到：在诊断报告与补救计划之间加入“本轮不补”字段之后，下一轮的属性掌握模式预测精度上升，则让层是一个独立对象，本章对；若观察到精度不变而只是失误参数的估计值改变，则让层已被这两个参数吸收，本章这一处错。

鉴别诊断与排除诊断（临床医学）。它说到这一步：诊断过程中会列出若干可能，逐一排除，最后留下一个。分离线在第六节那张表的两个格上——排除处理的是判定为不成立的，让层是判定为成立而不动的。两者在账面上都空，但后果相反。若观察到：一个被搁置的层在下次判定中被重新认定成立并被处置，它是让层；若它在下次被判定为从来就不成立，它是被排除的候选。这条观察不需要任何新工具，只需要连续两次的记录。

共病与多重诊断（现行疾病分类体系）。它说到这一步：一个人可以同时有多个诊断，编码体系提供了并列的位置。分离线在于共病是并列的，让层是有序的——让层结构自带一个“这一次动哪个”的指定，共病编码不含这个信息。若观察到：在共病编码被完整要求的制度下，主诊断的分布比只要求主诊断的制度下更集中，则医生确实在把有序的层压回单一诊断，本章对；若两种制度下主诊断分布无差异，则共病编码已经承担了分层的功能，本章这一处被吸收。这一条可在同一国家编码规则修订前后直接查。

技术债（Cunningham 1992，软件工程）。把本章的命题剥掉全部中医名词，只留下形式，它是这样的：一次判定同时产生一个处置对象和一组被认可而搁置的对象，后者不进入记录，因而在交接处丢失。这个纯形式的结构在上游领域早有名字，技术债是最像的一个。它说到这一步：为了赶进度而做的次优实现会积累代价，日后要还。分离线在于技术债是做了一个次优的东西并留下代价，

让层是没做而对象仍在——技术债的对象是已经被造出来的坏东西，让层的对象是还没被动过的真实层。更重要的是，技术债有账：缺陷跟踪系统里有 deferred 状态。若观察到：把技术债的记账形态（状态、负责人、目标版本）移植到中医病历后，复诊时该层被重新提及的比率显著上升，则两者形式同构而中医缺账，本章对；若观察到移植之后无变化，则中医的这个功能已由别处承担，本章的空缺主张为伪。

延期维护欠账（设施与基础设施管理，deferred maintenance backlog）。这是比技术债更贴的一位，也是本章必须正面承认的一位。它说到这一步：一项设施的维护需求被识别、被认可、而本年度不予执行，这部分被计入一个有正式会计科目的欠账，逐年累计，进入财务报告与审计。分离线在于这一位其实已经把本章要的字段造出来了——识别、认可、不执行、入账、跨期累计、可审计，六件齐全。本章因此不能声称这个字段无人想到过；本章能声称的只是：在临床判定、教育诊断和艺术鉴定这三处，这个已经成熟的记账形态没有被移植过。这一条对本章是削弱，本章接受它，并把它当作最好的现成模板——第十七节的赌注正是照它的形态写的。

保护基与无保护基合成之争（Hoffmann 2006；Baran 等 2007，化学）。它说到这一步：保护基步骤不增加分子复杂度，只增加步数，是账面损耗，理想的合成应当消除它。分离线在于同一份材料，两边给出相反的评价——化学界判保护基为浪费，本章判它为“这一次不动它”唯一被物质化的记录。这条分离线的锋利之处在于它不是“我比它多说了一点”，而是对同一个对象的两种相反读法。若观察到：无保护基路线在同类目标分子上系统性地牺牲了区域选择性或立体选择性，则保护基承载的是不可消除的分层信息，本章对；若观察到无保护基路线在选择性上无系统损失，则保护基确实只是损耗，本章借它做的论证要重做。

正式化的脱耦（Meyer & Rowan 1977，组织社会学）。本章第六节的签名三——越规范化越严重——是一个反转模板，而反转模板往往早有名字。最贴的一个是正式结构与实际活动的脱耦：组织采纳正式规则以获得合法性，而实际操作另走一套。分离线在于脱耦说的是两套并存且双方都知道，本章说的是正式结构里根本没有这个对象的位置。脱耦的前提是那件事在正式结构里有一个（被绕过的）位置；让层在正式结构里连位置都没有。若观察到：临床工作者能够指出病历系统中某个字段“实际上是用来记这个的”，则这是脱耦，本章的缺位主张

为伪；若访谈中工作者一致回答“没地方记”，则是缺位，本章对。这一条只需要一次结构化访谈。

体病相关与体质可调（中医体质学）。它说到这一步：体质与疾病易感性相关，体质相对稳定而可以调整。分离线在于**它把稳定性当作体质自身的性质，本章把一部分稳定性归给处置结构。**第九节之三已经给出了两组对照的设计。若轮转处置组的体质分型稳定性显著低于固定处置组而临床结局无系统差异，则本章对；若两组无差异，则稳定性确由体质自身给出，本章这一处错。

十条之中，最近的一位是延期维护欠账——它已经把本章要的记账形态完整地造出来了。本章之所以仍不被它吸收，是因为它处理的是一个**已识别需求在时间上的推迟**，而本章处理的是一次判定内部的分层与授权：延期维护里“要修什么”是先定的，推迟的只是时机；本章主张“是什么”这个判定本身就由“这一次动哪一层”切出来。这一步是延期维护不需要走的，因为设施的病害清单不依赖于本年度的预算而改变形状。

十一之一·这些说法共同看不见的那一块

把上面十位逐个问一句“它把什么当作给定，因此结构性地看不见什么”，答案有一个共同的形状。

兼证把“记下来的就是全部认定”当作给定，因此看不见“认定了而不记”这一类；一旦承认它，证候流行病学的患病率估计全部要重算，因为分母错了。鉴别诊断把“判定的输出是候选集合里的一个元素”当作给定，因此看不见“输出是一个带优先级的分层加一个授权”；承认它，鉴别诊断的整个决策树结构要重画。认知诊断模型把“诊断与补救是两个独立环节”当作给定，因此看不见“属性的粒度是按可教学单元切的”——而这一点它自己在设计文档里承认，却不允许它反向影响对属性的心理测量学解释；一旦承认这个反向影响，属性就不再是学生身上的东西，模型的解释基础要重建。古德曼把“同一性判据可以与使用意图无关地给出”当作给定，因此看不见“记谱法本身是按可执行性切的”；承认它，可记谱与不可记谱的分界就不再是艺术门类的性质，而是执行技术的历史状态。

四条盲区背后是同一块地：

一个类别的边界，是对象自身的分节，不是处置手段的投影。

这块地比第三节那一条更深。第三节推翻的是“判定不含处置授权”；这一条说的是“判定的**粒度**不由处置能力决定”。互相批评的各派都站在这块地上，因此谁也回头看不见它。本章这一次只掀开了它的一角——只处理了同一粒度下的分层，没有处理粒度本身如何被处置手段决定。这一角留在第十六节交代。

十二·与本系列各章的分界

本系列此前已有数篇处理相邻问题的论文，其中一篇与本章使用了完全相同的三个学科。本系列篇目是彼此最近的读者，也是彼此最容易被混同的对象，因此必须逐篇指名并给出可分辨的观察。

《撑平》（《治未病立账学》第一章）（何谓未病，同样取材于教育、化学、艺术这三处）。那一篇处理的是：一个读数持续正常，而这个正常由持续不断的双向抵消维持，被抵消的那份维持工作在账本上不设字段；它的读数是它持率。分离线在于**对象动没动**：那一篇里对象一直在被外力压回原位，账上看不见的是“谁在压”；本章里对象没有被压，账上看不见的是“这一次没打算动的是哪几层”。同一份“治疗后症状部分改善”的病历，那一篇要查的是**维持动作日志——是不是有人在持续代偿**；本章要查的是**本次处置针对的层数——是不是本来就只打算动一层**。两条查法指向不同的档案，因此两篇不可互相替代。更重要的是两者可以叠加：一个他持率高、同时让层比也高的病例，其疗效读数的可信度最低——因为它既有看不见的代偿，又有看不见的搁置。这两个数在数学上独立，在临床上却常常同现。

《叠裕》（《治未病立账学》第三章）（余量按单位分账）。那一篇说的是各单位全部合格，而这些合格依据同一份不可分割的储备重复整份计入，无账做过加总；它的读数是叠计比。分离线在方向：**叠计比看的是分母被横向重复使用，让层比看的是分子只覆盖了分母的一部分**。一个是同一份余量被多个单位各自声明，一个是同一个对象的多层只有一层被动。两者数学上互补：叠计比大于一说明分母被高估，让层比大于零说明分子被低估。同一份“四科全按指南规范用药”的模范档案，那一篇预测叠计比可达三，本章预测让层比接近零——因为各科都在动自己那一层，谁也没有让。这是两篇给出相反读数的一个具体场合。

《功债》（《治未病立账学》第九章）（措施退役，没有人负责减一层）。这是本系列与本章最近的一篇。它说的是加上去的措施无人负责撤除，残留的加法逐层累积。分离线在于**加与不加**：功债的对象是**已经施加而该撤未撤的措施**，

让层的对象是未曾施加而对象仍在的病机层。一个是残留的加法，一个是缺席的加法。同一份长期用药记录，《功债》要查的是“哪几味药本可停而没停”，本章要查的是“哪几层本次未动而下次也没动”。两者可以同时存在于同一份记录，且互相加重：让层多的病人更容易被加方，加方之后更难退役。

《阙前距》（《治未病立账学》第二章）（判定线的移动）。那一篇处理的是对象与判定线之间距离的两端只有一端被记账。分离线在于本章不动判定线——本章假定判定标准不变，只问在同一标准下被认定成立的层里动了几层。同一份连续二十年正常的档案，那一篇查判定线的版本历史，本章查每一次处置针对的层数。两个字段互不覆盖。

《悬量》（《治未病立账学》第八章）《漏汰》（《治未病立账学》第六章）《欠返》（《治未病立账学》第五章）（治未病三题）。这三篇分别处理症状的定位功能、筛选中的漏检、以及未完成的返程。与本章的共同分界很简单：它们处理的都是单层内部的读数结构，本章处理的是层与层之间的选择结构。若一个现象在只有一层的情形下仍然出现，它属于那三篇；若它只在两层以上时出现，它属于本章。这条判据可以在任何具体案例上直接执行。

《弃录》（本系列单篇）（被排除的候选物）。第六节的四格表已经给出分界：那一篇是第④格（不成立而不动），本章是第②格（成立而不动）。两篇合起来才把“不出现在账上的东西”这一片补齐，但两者的补救方向相反——那一篇要求记录被淘汰的理由，本章要求记录被保留而未动的对象。若一个未出现在处置里的层，在下一一次被处置了，它归本章；若它在下一一次被判为从来不成立，它归那一篇。

本系列这七篇共用一个未言明的做法：都在找各领域账本上被当作垃圾、余数或噪声的那一格。这是它们共同的强处，也是它们共同的弱点——第十六节会交代这个弱点的具体形状。

十三·三处反过来削弱本章的地方

一个跨领域的判断，若三个来源全都顺着它说，那它多半只是在这三处各挑了一句合用的话。真正的兑现，必须包含来源反过来限制它的地方。本章有三处，其中第一处动的是适用范围与价值排序。

第一处，化学削掉了本章最想主张的那句话。本章原本会写下这样一句更强的话：分层是判定的普遍结构，任何判定都必然携带一个处置授权，因而任何领

域都应当为让层建账。无保护基合成这一支主张不允许本章这样写。它的论证是硬的：当试剂的选择性足够高、或目标分子的官能团差异足够大时，保护基是可以被彻底消除的，而且消除之后路线更短、损耗更小。也就是说，**分层不是判定的普遍结构，它是当处置手段的选择性低于对象的分节精度时出现的代偿。**手段足够精准时，让层比趋于零，而这不是记账缺失，是真的没有让层。

本章因此把适用范围缩到这样一类系统：**对象的分节精度高于处置手段的选择性，因而一次处置不可能只作用于一层而不牵动其余。**中医的方药显然在这一类里——一张方作用于全身，不存在只针对某一层的手术刀。教育的补救教学在这一类里。文物修复在这一类里。而现代药理学里高选择性的靶向药，可能正在离开这一类；若一种药只作用于一个明确的靶点而不影响其余，那么关于它的判定确实可以不带分层。

这一处削弱同时改掉了价值排序：本章不再主张让层比越低越好，也不主张越高越好。**它主张的只是这个数应当被记下来。**

第二处，教育测量削掉了本章关于盲区范围的主张。本章原本想说：所有这些领域都看不见“判定的粒度由处置能力决定”。教育测量不允许这样说——Q矩阵的属性划分按可教学单元切，这是公开的设计原则，写在方法学文献里，不是盲区。教育测量清楚地知道它的属性是按干预单元切的。

它看不见的是另一件更窄的事：**同一份诊断里被认定为未掌握而本轮不予补救的那些，没有被当作一个对象。**本章的空缺主张因此要缩小——不是“这个领域看不见粒度问题”，而是“这个领域看得见粒度问题，却仍然没有为分母减分子那一块设字段”。这个缩小让本章的主张更窄，也更难被反驳。

第三处，文物与艺术品保存削掉了本章关于“不设字段”的证据强度。本章原本会写：这个差在所有考察的领域里都不存在于任何记录中。文物与艺术品保存不允许——现状病害报告与修复处理报告并存，差额在原则上可以直接算出。设施管理的延期维护欠账更进一步，它已经有正式会计科目、跨期累计、可审计。

本章因此把主张降为一个更弱但更准确的形式：**这个差在若干领域里是可算的，但它没有名字，没有人算，不参与评价，也不被下一次判定继承。**空缺不在数据层，在**对象层**——材料在那里，而它不是一样东西。这个降级使本章的第五节之一必须重写，本章已按此重写。

三处削弱之后剩下的主张，比本章最初打算写的窄得多。这是应该的。

十四·不适用的边界

本章的判断在下列情形不适用，或需要重大修改：

处置手段的选择性高于对象分节精度时。 见上一节第一处。此时让层可能真的不存在，而不是没被记。

单层情形。 若一次判定只认定了一层，让层比为零，本章的全部结构退化为平凡。这不是缺陷，是适用条件——但它意味着本章在急症、单一明确病因的情形下几乎没有内容。

判定与处置由不同的人做出时。 本章假定辨证与处方是同一个人在同一次里完成的。在分工的系统里（诊断科室与治疗科室分离），处置授权可能确实不在判定里，本章的核心命题在那里要重新论证。

处置不可分的情形。 若一次处置在物理上无法只作用于部分层——例如一次不可分割的手术、一次全身照射——那么“这一次动哪一层”这个问题没有意义。让层需要一个可以选择作用范围的手段作为前提。

记录本身即处置的情形。 在心理治疗与某些教育场合，说出“我这次先不处理这一层”这个动作本身就是一种干预，会改变对象。此时为让层建账不是中性的记录，它有治疗后果。本章不处理这种情形，也不建议在这种情形下机械地套用第十七节的赌注条款。

跨文化的诊断体系。 本章全部材料取自一个记录传统。它是否适用于其他诊断传统，本章没有证据。

十五·可以让本章翻车的观测

一个判断若给不出会让自己翻车的观测，它就还不是一个判断。本节按三层给出，因为本章的主张可以被分层引用。

第一层是让层这个构念本身，最强，也最容易倒。**第二层是那张四格表**，作为分析工具，即使构念被推翻仍可能有用。**第三层是一条不依赖前两层的经验主张**：证候积分法的分母未做分层调整，因而系统性低估分层治疗方案的第一阶段疗效。这一层最容易检验，而且检验它不需要接受本章的任何概念。

十五之一·最强的竞争解释，以及排除它的条件

必须先写出那个最朴素、最难反驳的说法，而不是挑一个好打的。它是：

其实不就是医生记录不全吗？

这个解释很强，因为它符合所有人的经验，而且它的处方很便宜——加强培训，加个自由文本框，就完了。

排除它的条件写成这个形状：

控制记录完整度（以同一诊次的完整录音转写为基准，比对结构化病历的信息保留率）之后，若“诊疗过程中被口头认定成立的层数与处置针对层数之差”与“该层在复诊时被重新提出的概率”之间不存在剂量—反应关系，则本判断为伪。

关键在于本章与竞争解释给出的处方不同，而这个不同是可以直接检验的：

- 若只是记录不全，那么**增加自由文本容量**就应当有效。
- 本章预测**增加自由文本容量无效**，而**增加一个带状态、可被下一次就诊自动带出的结构化字段有效**。理由是问题不在容量，在于没有任何字段承担跨次继承。

这两个处方可以在同一批门诊里并行试，成本很低。

十五之二·样本纪律

凡涉及剂量—反应关系的检验，样本至少六个单元，且在关键混淆变量上同质。本章明确排除跨国或跨地区的两两比较作为证伪检验——两个国家在制度、语言、记录传统、医患关系上处处不同，所谓“控制掉某变量”在方法上站不住，那样的比较只能作启发式说明。

可用的同质设计有：同一家中医院的六个以上科室；同一省内多家三级中医院的同一科室；同一教学医院不同年资的多组医师；同一个电子病历厂商的多家客户医院。自变量（是否设置让层字段、字段是否可继承）在这些单元间存在真实变异，而混淆变量大体同质。

十五之三·其余可以让本章翻车的观测

一·若在辨证记录中增加“本次认定成立的全部层”一项后，评价者间一致性没有上升，则本章关于“分歧主要在排序而非在认定”的主张为伪。

二·若“本次认定成立的全部层”的一致性上升幅度不显著大于“本次处置针对层”的一致性上升幅度，同上。

三·若轮转处置组与固定处置组的体质分型稳定性无差异，则第九节之三关于体质是让层沉淀的主张为伪。

四·若在诊断报告与补救计划之间加入“本轮不补”字段后，下一轮属性掌握模式的预测精度不变，则让层在教育测量里已被失误与猜测两个参数吸收。

五·若无保护基路线在同类目标分子上的区域选择性与立体选择性无系统损失，则保护基确实只是账面损耗，本章借它做的推翻论证要重做。

六·若结构化访谈中临床工作者能一致指出病历系统中某个现有字段“实际上就是用来记这个的”，则本章的缺位主张为伪，此处是脱耦而非缺位。

七（正向读数，且必须写成顺序而非相关）·若让层字段被引入，那么“该层在复诊中被重新提出”的比率上升，应当早于“复诊处置针对层数”的上升。认定被继承在先，处置跟上在后。若观察到两者同时上升，或处置先动，则本章关于“缺的是继承而非意愿”的机制解释为伪。

八（低成本可实做，且所需数据已经存在）·取任意一家中医院近三年的门诊病历与处方，由方剂学判读每张方针对的层，与病历“证候”栏比对。若绝大多数处方针对的层数等于病历记录的证名数，则让层现象不存在。这个检验不需要采集任何新数据。

若本章被证伪，学到的东西是明确的：那意味着中医辨证的输出确实是一个完整描述而非一个带授权的分层，因而证候标准化那条路的方向是对的，只是精度不够。这是一个有价值的结论，而且它会把资源引向完全不同的地方。

十五之四·本章解释不了的两个洞

第一个洞：让层放多久会变成另一样东西？本章说体质是让层的沉淀，但说不出这个沉淀需要多少次、多长时间，也说不出它是渐变还是有阈值。这一层本章认定成立而这一次不动。

第二个洞：换医时让层由谁持有？本章说它在交接处丢失，但没有说清一个理想的交接应当把它交给谁——交给病历、交给病人、还是交给某个跨机构的记录。这一层同样认定成立而不动。

十六·本章自己的让层比

按本章自己的判据给本章定位，得到的结果不好看，而且必须说出后果。

本章这一次认定成立的层至少有四层：让层不设字段（本次处置的那一层）、让层的时间性即沉淀为体质的机制（第十五节之四第一个洞）、让层在交接时的持有者（第二个洞）、以及判定粒度本身如何被处置手段决定（第十一节之一末尾掀开一角而未处理的那一块）。本次只动了第一层。

本章自己的让层比是四分之三。

后果有三个，逐一说清。

后果一：本章正落在自己指认的那一格里。第六节说那一格的三条签名是短期指标改善、失效不产生信号、越正规化越严重。本章完全符合：只动一层使论证集中而锋利（短期好看）；其余三层若无人接手，不会有任何人发现（无信号）；而本章若被引用、被规范化为一个可操作的字段方案，那三层会被更彻底地挤出视野——因为一旦有了“让层字段”这个正式名目，人们会认为这件事已经被处理了。**本章的成功，正是它自己描述的那种自我加固的最好例子。**

后果二：本章与本系列诸篇共用一份余量，而这份余量没有被加总过。本系列此前七篇有一个共同做法——都在找各领域账本上被当作垃圾、余数或噪声的那一格。这个做法是本系列全部篇目共享的一份洞察储备，而每一篇都把它整份计入了自己的贡献。按本系列《叠裕》一篇自己的读数，**本系列这一系列的叠计比至少是八。**这意味着：若“寻找账本外的对象”这个做法本身被证明是一种系统性的错觉——例如被证明凡是各领域都不记的东西，恰恰是因为它在实践中真的不重要——那么本系列这些篇目**不会依次失效，而会同时失效。**本章不打算掩饰这一点，也提不出办法把这份储备分账。

后果三：本章的处方一旦被采纳，最省事的实现方式会摧毁它要保护的东西。这是本章的反噬预言，而且它没有出口。

设想让层比进入了考核。降低它最省事的办法不是多动几层，而是在**认定阶段就少认几层**——把那几层不写成“成立”，于是分母缩小，比值好看，而真实的分层判断退回医生脑内，比现在更不可见。这个做法在任何审计下都合法，因为“哪几层成立”本身就是医生的判断，外部无从核验。任何要求“必须完整列出全部成立的层”的规定，都可以被“我认为只成立一层”合法地满足。

有没有出口？本章找过三条，三条都不通。第一条是要求列出鉴别过程——但鉴别过程记的是被排除的（第④格），不是被认定而不动的（第②格）。第二条是用处方反推层数——但反推只能得到被处置的层，得不到被让开的层。第三

条是不把它用于考核而只用于研究——但一个数一旦存在于系统里，它进入考核只是时间问题。

我看不到出口。本章能做的只有第八节之二那三条限制，而三条限制全部依赖执行者的自我约束，这恰恰是本章自己论证过的最不可靠的东西。

十七·一条写死日期的赌注

一个判断若不肯把自己压在一个可以到期兑现的事实上，它就还只是一种说法。本章压这一条：

在二〇三五年十二月三十一日之前，中医电子病历的国家级或学会级结构化标准（或《中医病历书写基本规范》的修订版）中，将出现一个独立于“兼证”的、可被下一次就诊自动带出的“本次未处置病机层”字段，且该字段的内容进入证候疗效评价的分母调整。

赌注若不写死什么不算命中，它就可以被任何模糊的进展宣布为兑现。因此以下五条写死：

一·只在“按语”“医嘱”“诊疗思路”等自由文本中体现的，不算命中。本章的主张是这个对象需要一个结构化字段，不是需要更多书写空间。

二·仅要求医生“说明治疗思路”或“填写辨证依据”而没有独立字段的，不算命中。说明治疗思路记的是为什么动这一层，不是没动哪几层。

三·字段存在但不进入疗效评价算式（证候积分仍按全部症状计入分母）的，不算命中。若不进入算式，这个字段只是一个不被读取的记录，第五节之一已经论证过这种记录等于不存在。

四·仅在科研专用病历、课题病例报告表或某单一机构的内部系统中出现，未进入常规诊疗规范的，不算命中。本章的主张针对的是记账制度，不是研究工具。

五·字段名虽异而实质仍是“兼证”（即填写后要求本次一并处置）的，不算命中。这一条最要紧：要求一并处置的字段，恰恰取消了让层这个对象本身。若最终落地的是这样一个字段，本章不但不算命中，本章的主张实际上被否定了——因为那意味着制度承认了这些层的存在，同时否认了“认定成立而本次不动”这件事的正当性。

这条赌注的形态是照着设施管理的延期维护欠账写的（第十一节）。那一处已经把六件做齐：识别、认可、不执行、入账、跨期累计、可审计。本章赌的是这套形态会被移植到临床判定上。移植失败也是一个结论——它会说明临床判定与设施维护在结构上有一处本章没有看见的根本不同。

十八·结论

本章从一道两千年的定义题出发：一个「证」到底是什么，它与病、与症、与体质的界在哪里。

本章的答案不是又一个更细的定义。本章认为困难来自一个共同假定：辨证是一次描述，先认清对象，再决定处置。三个互不通约的传统——作品同一性的记谱法判据、化学的机理判据、教育测量的情境判据——在“该由什么裁定”上势不两立，在这个假定上完全一致。而推翻它的材料就在其中一家自己手里：在化学里，一个分子有几个反应位点，只有相对于一个给定试剂才有答案；保护基是“这一次不动它”被物质化的唯一形态。

由此得到的主张是：一个「证」是一次带处置授权的分层判定。它在同时被认定成立的若干层里指定本次动手的那一层，并把其余各层认定为成立而不动。那些被让开的层，在病历、证候标准、疗效评价这三套账上都不设字段。

两根轴给出四个格：成立而动的是本证，不成立而动的是误治，不成立而不动的是正确排除，**成立而不动的是让层**。第④格无后果，第一第③格已有制度，只有第②格既有后果又无位置——而它通过全部形式审查，短期指标上表现为改善，失效不产生信号，且越规范化越严重。

一句不含情态词的问话把它落地：**这一次的处置针对的项，与这一次判定认定成立的项，数目是否相同？不同的那些记在哪个字段？**由此得到一个在中医、化学、教育测量、文物与艺术品保存四处量纲不同而比率相同的读数——让层比。它与既有主指标不但正交，在若干情形下结构性地反向：收率最高的合成路线保护基最多，证候积分最难看的病例可能是分层最清楚的那一个。

四组分界因此可以重做：症是判定的输入项，证是本次判定的分子，病是跨次被反复重新认定的层集，**而体质是长期居于分母而从不进入分子的那些层的沉淀**。证与体质在单次诊察里原则上不可分，不是因为技艺不精，是因为它们在单次诊察里本来就是同一样东西。分界只能出现在跨次的账上，而那本账现在还没有人在记。

三处削弱把这个主张收窄了：处置手段足够精准时分层不必要；教育测量看得见粒度问题；文物与艺术品保存与设施管理已经把这个字段的形态造出来了，只是没有被移植到临床。收窄之后剩下的主张是：**在处置手段的选择性低于对象分节精度的那一类系统里，判定必然携带一个分层授权，而被授权不动的那些层，需要一个字段。**

本章自己的让层比是四分之三，本章正落在自己指认的那一格里。本章与本系列诸篇共用一份未经分账的洞察储备，因而它们会同时失效而非依次失效。而本章的处方一旦进入考核，最省事的达标法是在认定阶段少认几层——对此我看不到出口。

最后那条赌注，压在二〇三五年十二月三十一日：一个独立于兼证、可被下一次就诊带出、且进入疗效评价分母调整的“本次未处置病机层”字段。五条不算命中的情形已经写死，其中最要紧的一条是：**要求填写后必须一并处置的字段，取消的正是这个对象本身。**

第二章 吞差

——一个「证」为什么是链上的一段而不是一个类别：论“判定为相同”这个结果所吸收的量为何在任何账上都不设字段

提 要 「证」是中医全部临床操作的枢纽，而它至今没有一条稳定的边界。本章不再问这条边界应当划在哪里，而问一个更前面的问题：“**同一个证**”这个关系，是不是一个可传递的关系？本章的答案是否定的，并由此提出一条主张：一个「证」不是一组症状共变背后的共同原因，不是某个尺度上粗粒化出来的宏观态，也不是给定观测配置下可被区分出的一个类，而是一条由相邻不可分辨关系串起来的链上的一段——它的边界不是它自身的性质，而是这条链在何处被登记过一次“改变”。推翻上述共同假定的材料来自心理学自己：心理物理学早已证明，“分辨不开”不是等价关系，相邻两两不可分辨的序列，两端可以明确分辨。由此得到一类在三套账上都不设字段的东西：在一串各自合法的“判同”里被逐次吸收、在任何一次记录中都不出现、而在跨次上可加的那些差异量，本章称为吞差。本章给出一句不含情态词的问话：把第一次记录与最近一次记录直接并排，不看中间，会不会判出同一个证？记录里登记过几次证型改变？由此得到一个跨领域量纲不同而比率相同的读数——吞差率。它与既有的辨证一致性指标结构性地反向：一份“二十次复诊证型始终一致”的模范档案，吞差率可以是一。本章并指出，中医语言里“同一个证”混用了两种不同的关系——相邻同一与跨段同一，这是两千年混乱的一个来源。

一·从“边界划在哪里”退一步

关于「证」的争论，绝大部分发生在同一个层面上：这个证与那个证的界限应当划在哪里，条目该怎么定，阈值该取多少，主症次症如何配比。近三十年的证候标准化工作，绝大部分精力花在把这条界限划得更细、更稳、更可复制。

这条路有一个隐含的前提，而这个前提从来没有被单独拿出来检查过：

“同一个证”是一个可传递的关系。

也就是说：若这一次与上一次是同一个证，上一次与再上一次也是同一个证，那么这一次与再上一次就是同一个证。这条推理在日常语言里天经地义，在任何分类学里都是分类得以成立的基础——一个类别若不能把它的成员两两归为同类，它就不是一个类别。

本章要论证的是：在临床判定里，这条推理不成立，而且它的不成立不是偶然的误差，是结构性的。

先看一个几乎每天都在发生的场景。一位慢性病患者每月复诊一次。第一次辨为脾气虚。第二次，医生看了，与上次比没有明显变化，仍记脾气虚。第三次、第四次，同样。二十四个月之后，病历上是一串整齐的“脾气虚”，中间从未登记过任何一次证型改变。

现在做一件从来没有人做的事：**把第一次的四诊记录与第二十四次的四诊记录抽出来，遮住日期，交给一位不知道它们来自同一个人的医生。**

他很可能判出两个不同的证。

这不是假设。它是相邻不可分辨关系的必然后果，而这个后果在心理学里被研究了一百多年。本章第四节会说清它的来源。这里先把问题的形状点出来：

中间的二十三次判定，每一次都是对的——相邻两次之间确实分辨不开，任何一位有经验的医生在那两次之间都会做出同样的判断。**而整体是错的**——首末两端确实不是同一个证。

一串正确的判定，累加出一个错误的记录。中间没有任何一次可以被追责，因为没有任何一次是错的。

这个结构在中医之外的领域里也有，而且有些领域已经解决了它。儿科不再只问“这次比上次长了没有”，而是把每一次身高体重放回一条跨年的百分位曲线上——**生长曲线这个发明，全部意义就在于对抗相邻比较的累积失效**。工业设备的状态监测也不再只与上一次读数比，还要与出厂基线比、与同型机队比。这两处都在明确地承认：**相邻比较会漏掉整体漂移，而漏掉的部分不会自己显形**。

中医没有这样一条曲线，也没有这样一个基线。证型的判定几乎全部是相邻比较——医生看的是“较前如何”。这不是疏忽，这是记录结构决定的：病历按时间流水书写，前一次就在眼前，首诊在二十四页之前。

本章由此提出的问题不是“证的边界该划在哪里”，而是：

每一次“判定为相同”，究竟吸收了多少？这个量记在什么地方？

答案是：不记在任何地方。而且它不是记得少，是不算一样东西。

本章的路线：第二节请三个互不通约的传统各出一条最强的答法；第三节找出它们底下那个谁都没说出口的共同假定；第四节从其中一个传统自己的材料里找出推翻它的那一件；第五节到第八节给出主张、两根轴、一句可以拿去问记录系统的问话，以及一个可以在几个领域并排比较的数；第九节回到题目，说

清“何谓一个证”；其余各节交代兑现、与相近说法的分界、削弱、边界、可以让本章翻车的观测，以及一条写死日期的赌注。

二·三条不肯让步的答法

“一样东西凭什么算是同一样东西”，这个问题有三个传统给出过强硬且互不相容的答案。本章取它们最强的版本。

二之一·第一条：看当下的共变结构就够了

这一条来自心理测量学，而且它是这个领域里最不肯让步的一支。

它的主张是：一个构念要算作一样真东西，它必须是那些指标共变的共同原因。若干个症状之所以一起出现，是因为背后有一个东西在同时驱动它们；这个东西就是构念。它有一条可以执行的判据——**局部独立性**：一旦把这个潜在的东西固定住，那些症状之间就应当不再相关。它们剩下的相关，全部由那个共同原因解释掉了。

这条判据的锋利之处在于它能把真东西和假东西分开。若把一组指标加个总分就叫一个构念，而给定总分之后指标之间仍然强相关，那么这个总分只是一个求和记号，不是一个实体。它可以用来排序，不能用来解释。

把这一条移到中医上，它对应的是这样一种读法：**脾气虚是一个真东西，当且仅当纳呆、便溏、乏力、脉缓这些表现之所以一起出现，是因为背后有一个共同的东西在驱动。判据是：控制住这个东西之后，这些表现彼此不再相关。**它不关心病人是怎么走到今天的，也不关心是谁在什么条件下看的。它只要求当下这一横截面的共变结构给出答案。

它的主张压成一句：**看当下的共变结构就够了。**

二之二·第二条：看它在哪个尺度上被组织起来才够

这一条来自物理学。

物理学在很长时间里也以为宏观态就是那么摆在那里的。二十世纪后半叶之后，这个想法被彻底改写了。一个系统的正确描述变量，取决于**你在哪个尺度上把哪些区别抹掉**。粗粒化的选择不同，有效自由度就不同；跨越了一个特征尺度，原来的变量甚至不再是变量。有效理论这一套做法的核心就在这里：低能标下的正确描述，与高能标下的正确描述，不是精粗之别，是**不同的东西**。

物理学对这件事有一个非常清醒的表述：一个宏观态是一大堆微观态的等价类，而这个等价类由你选定的宏观变量集合定义。换一组宏观变量，等价类就重划。

更值得注意的是，物理学**为被抹掉的那部分建了账**。熵就是这笔账——它数的正是“在这个粗粒化下，有多少微观区别被当作了同一件事”。这是三个传统里唯一一处，把“被判为相同的那些差异”变成了一个中心量。第十三节会说明这一处如何反过来削弱本章。

把这一条移到中医上，它对应的是这样一种读法：脾气虚不是一个实体，它是一个粗粒化的名字——在四诊这个分辨率上，一大批不同的身体状态被归为同一件事。换一个分辨率（加上生化指标、加上影像、加上代谢组），这个归类就要重划，而且新旧两套之间不是精粗之别。

它的主张压成一句：**看它在哪个尺度上被组织成这样才够。**

二之三·第三条：看它此刻纠缠着什么观测配置才够

这一条来自工程学的故障诊断理论。

工程学问过一个非常干脆的问题：一个故障是不是“可诊断的”？它给出的答案是一个联合性质——**可诊断性不是故障的性质，是故障与传感器配置这一对的性质**。同一个物理故障，在一组传感器下可以在有限步内被唯一识别，在另一组传感器下则与另一个故障永远不可区分。诊断理论里有一个正式的概念叫不可区分故障对：这两个故障在所有可能的观测序列上产生完全相同的输出。

工程学由此得出一条几乎冒犯常识的结论：**问“这台机器出的是哪个故障”，在传感器配置给定之前，是一个没有答案的问题**。而传感器配置是可以花钱改变的——加一个测点，两个原本不可区分的故障就分开了。

把这一条移到中医上，它对应的是这样一种读法：一个证是什么，取决于四诊这套采集手段能分辨什么。望闻问切是一组特定的传感器，它有它的分辨极限。在这套配置下，若干个不同的病理状态是不可区分的；换一套配置，它们就分开了。**问“这个人到底是哪个证”，在采集手段给定之前，没有答案。**

它的主张压成一句：**看它此刻纠缠着什么观测配置才够。**

二之四·三条为什么不能同时成立

第一条与第二条互相取消。心理测量学要求构念是共同原因，有因果实在性；物理学说宏观态只是一个等价类的名字，它不是任何东西的原因——宏观态不引起微观态，方向是反的。若前者对，那么任何一个不能通过局部独立性检验的证都该被取消资格；若后者对，那么局部独立性这条要求本身就问错了问题，因为一个粗粒化的名字本来就没有义务当原因。

第二条与第三条是最值钱的一对，因为它们构成了根基与病灶的关系。物理学每天在做的那件事——**选定一组宏观变量、抹掉其余区别**——正是工程学诊断理论判为问题所在的那件事：在工程学看来，“抹掉哪些区别”从来不是理论内部的选择，它是传感器配置的后果，是一个可以花钱改变的工程事实，把它当作理论选择就是把工程问题误认成物理问题。两边不是意见不同，一边的地基是另一边的病灶。

第一条与第三条互相取消。心理测量学的判据是内在的——局部独立性是数据内部的性质，给定一份数据就能算。工程学说这类问题在原则上不能内在地回答：换一组传感器，同一批底层状态产生的数据就变了，共变结构跟着变，局部独立性检验的结论也跟着变。若后者对，那么任何一次局部独立性检验的结论都只是关于“这套采集手段”的结论，不是关于那个构念的结论。

三条不能靠判谁对来化解，因为它们不在同一个语域里作答。要处理它们，只能往下挖。

三·三条答法共同站着的那块地

三条答法争的是同一件事：“**这两个情形算不算同一个证**”该由什么裁定。一条说由当下的共变结构裁，一条说由粗粒化尺度裁，一条说由观测配置裁。

它们争得很凶，而有一样东西三边谁也没有提出来过，因为在它们看来这根本不需要提：

给定一个分辨标准之后，“**算不算同一个**”这个问题有确定答案；而这个答案构成一个可传递的关系——凡是被判为相同的，可以被归进同一个类。

这句话念给三边听，三边都会说这还用说吗。

心理测量学那一边：给定一个潜变量模型，两个个体是不是属于同一个潜类，模型会给出后验概率。潜类分析的整个框架建立在潜类是**互斥且穷尽**的分割上——这正是一个等价关系。

物理学那一边：给定一个粗粒化，两个微观态是不是属于同一个宏观态，答案是确定的。**等价类这个词本身就写在定义里**：宏观态就是微观态的等价类，而等价类要求自反、对称、传递，三者缺一不可。

工程学那一边：给定一组传感器，两个故障是不是可区分，答案由可诊断性分析给出。不可区分关系被当作一个把故障集合分成若干**不可区分类**的关系——同样是等价关系。诊断理论里这些类被直接叫作 indistinguishable classes。

三边在“该由谁裁”上势不两立，在“裁完之后得到的是一个可以归类的关系”这一点上完全一致。而中医的证候研究整个建立在同一块地上：证候标准把证切成互斥的类型，证素辨证把它拆成可组合的要素，隐结构分析给出潜在的类别划分。**这些方法的输出全部是分割，而分割预设传递性。**

这块地是本章要挖开的那一块。挖它不能从外面搬一个理由——那样只是加入争论，成为第四种立场，三边可以各自退回自己的阵地。要挖开它，必须用三边之一自己已经掌握、已经发表、只是从来不住这个方向用的材料。

四·推翻它的那件材料，来自心理学自己

心理学有一个部门，一百五十年来一直在测同一件事：**人能分辨多小的差别。**

心理物理学的基本工具是最小可觉差。给一个刺激，慢慢加一点点，问被试“和刚才一样吗”。加到某个量，被试开始说“不一样”。这个量就是这个刺激水平上的最小可觉差。

这套工作产出了大量定律和方法。但它同时产出了一个结论，这个结论在心理物理学内部人尽皆知，而它的分量从来没有被送到别的地方去：

“分辨不开”不是一个等价关系。它不传递。

具体形状是这样：取一系列刺激，相邻两个的差都小于最小可觉差。于是相邻两两之间，被试一律报告“一样”。而把这一列的第一个与最后一个直接放在一起，被试明确报告“不一样”。

颜色是最经典的例子。一排色片，每一片与它旁边那一片肉眼分辨不开，而这一排的两端是两种明确不同的颜色。灰度、音高、重量、亮度，全都如此。这不是被试不专心，不是方法粗糙，不是样本量不够。**它是有阈限的分辨系统的必然性质**——只要分辨能力有一个大于零的下限，不可分辨关系就一定不传递。

数学上这件事早已被正式处理过。这类关系不构成等价关系，它构成的是另一种结构：**相邻可容忍，整体有序**。要描述它需要放弃传递性，保留一个更弱的性质。心理学与决策研究里为此专门发展了一套形式工具。

现在请注意这件材料在心理学自己领域里回答的是哪个问题：**它回答的是“阈限有多大”**。它是关于感觉系统灵敏度的知识，用来标定仪器、设计实验、编制量表。

它从来没有被拿来当作关于“构念同一性”的证据。

而它就是那件证据，而且它硬到没有讨论余地。心理测量学主张一个构念的实在性由局部独立性这条内在判据裁定，而局部独立性预设了潜类是一个分割，分割预设传递性。**心理学的另一个部门，用一百五十年的实验证明了：任何有阈限的判定系统，它给出的“相同”关系都不传递。**

而临床辨证是一个有阈限的判定系统。四诊有分辨极限，这一点没有人会否认。

推翻之后，问题就变了。原来的问题是“证的边界该由哪一维裁定”，现在的问题是：

如果“判定为相同”这个关系不传递，那么一串“判定为相同”里被逐次吸收掉的那些差异，去了哪里？

五·主张：吞差

本章的主张是一句三重否定加一个新东西：

一个「证」不是一组症状共变背后的共同原因，不是某个尺度上粗粒化出来的宏观态，也不是给定观测配置下可被区分出的一个类，而是一条由相邻不可分辨关系串起来的链上的一段——它的边界不是它自身的性质，而是这条链在何处被登记过一次“改变”。

由此得到一类东西：

在一串各自合法的“判定为相同”里被逐次吸收、在任何一次记录中都不出现、而在跨次上可加的那些差异量。

本章称之为吞差。

三个否定各对应一边。第一个否定的是共同原因判据，第二个否定的是粗粒化判据，第三个否定的是观测配置判据。三边都不会同意这条主张——三边都会说，你把一个测量精度问题说成了本体问题。但三边的证据合起来只能得出它。

五之一·吞差不是什么

吞差不是误差。误差是判定与真值之间的偏离，它有方向、有分布、可以被平均掉。吞差不是偏离——每一次判定都是对的。二十三次判同，二十三次都正确。吞差是正确判定的产物，不是错误判定的产物。这一条至关重要：任何试图靠提高判定准确率来消除它的方案都不会有效果，因为它不来自不准确。

吞差不是漂移。漂移说的是被测对象或测量基准在缓慢移动。这个说法本身没错，但它把问题定位在“变化太慢所以没察觉”。吞差的定位不同：**即使变化并不慢，只要相邻两次的间隔小于分辨阈，吞差照样发生。**关键变量不是速度，是采样间隔与分辨阈的比值。缩短复诊间隔会加重吞差，而不是减轻——这是一条与常识相反的推论，第十五节把它列为可以让本章翻车的观测之一。

吞差不是记录惯性。电子病历里的复制粘贴、“较前无变化”的套话、上一次的诊断被自动带出，这些都是真实存在的问题，医学信息学有专门文献。但它们是书写行为问题。吞差是判定结构问题：**即使每一次都完全独立地重新辨证、不看上一次的记录，吞差照样发生，**因为相邻两个时点的状态确实分辨不开。这给出一条可以直接检验的分离线，见第十一节。

吞差不是模糊。模糊说的是边界地带的隶属度介于零一之间。模糊集处理的是“这一个属于该类的程度”，它仍然是一个关于单次归属的说法。吞差是一个关于一串归属之间关系的说法，而且它要求可加。模糊隶属度不可加。

五之二·它为什么是一类原来不存在的东西

有一个检验很硬：删掉这个读数之后，它还在不在？

若一样东西删掉读数之后照样存在、只是不好测，提出读数的人做的是测量改进。若删掉之后它压根不算一样东西，提出读数的人做的是造出一类东西。

吞差属于后者。三份账可以逐一去查。

病历这一份账：中医电子病历的结构化字段记录每一次的证型判定。若这一次与上一次判定相同，记录上是同一个证名——**没有任何字段的语义是“这一次的判同吸收了多少”。**判同在记录上是一个零，而零不携带任何量。

一致性评价这一份账：辨证一致性用一致率或 Kappa 来算。这类指标的结构是数“判同的次数”与“判异的次数”，**判同一律被当作零信息**。两位医生对同一份材料都判“脾气虚”，记为一致，记完为止。至于这两个“脾气虚”是不是指向同一批身体状态，指标不问，也无法问。

疗效评价这一份账：证候疗效按治疗前后的积分差算。基线是首诊。若病情在治疗期内缓慢漂移而证型判定始终不变，漂移的部分会被计入积分变化——但它被读作“疗效”或“无效”，**不会被读作“这段时间里判定吸收了一个变化”**。

三份账都不设字段。所以在这三份账里，吞差**不是记录得少，是不算一样东西**。删掉本章这个读数，它不会变成难以测量的东西，它会回到它现在所在的地方——不存在。

五之三·一个立刻可以做的推论

如果吞差是真的，有一件事应当已经在发生：

在辨证一致性最高的那些临床单位里，首末直接比较的不一致率也应当最高。

一致性与吞差率不是同向的两个好指标，它们结构性反向。一个每次复诊都与上次保持一致的单位，正是吞差累积最快的单位。

这个预测不需要新数据。它只需要把已有的病历按一个从来没有人用过的方式切一次。

六·两根轴与四个格

第一根轴是**相邻两次是否被判为相同**。第二根轴是**首末两端直接比较是否可分辨**。

	首末直接比较：可分辨	首末直接比较：不可分辨
相邻各次判同	② 吞差——每一次判定都对，整体记录是错的	① 真稳定——确实没变，记录忠实
相邻有判异	③ 正常追踪——变化被登记了，记录跟上了	④ 过度反应——记录里变了而整体没变

四格都不是虚设。

第①格是理想状态：病情稳定，记录忠实，两端一致。

第③格是记录系统正常工作的样子：变化发生了，被登记了，首末一致地反映出来。

第④格是一眼可见的坏，而且现有制度已经在惩罚它——辨证不稳、证型频繁变动、一致性指标低。质控条目里有它的位置。

第②格是本章要打的那一格。

六之一·为什么要打的是第②格

选定要打的那一格有一条判据：**它必须是全部形式审查都能通过的那一格。**一眼可见的坏不需要任何新框架。

第②格通过全部形式审查。病历完整，每一次辨证都有据，前后记录一致，证型稳定，一致性指标漂亮。翻遍现行的病历质控条目，没有一条会因为第②格而扣分——因为质控从来只查单次记录的完整性与合规性，从不做首末的直接比对。

它还满足另外三条签名。

签名一：它在短期指标上表现为改善。证型稳定被读作病情稳定。辨证一致性高被读作医疗质量好。评价者一致性是证候标准化研究的核心指标之一，而吞差正是把这个指标推高的最有效方式——只要每次都与上一次保持一致，一致性必然漂亮。

签名二：它的失效不产生任何信号。二十三次判定没有一次是错的，因此没有任何一次可以被追责，也没有任何一次会触发复核。当首末的差异终于大到无法忽视时，它会以另一种形式出现——病人被认为“突然加重”、“病情转变”、“新出现了某个证”。而实际上它是一路走过来的，只是每一步都被吸收了。这个误读的方向是固定的：**总是被读成一次突变，从不被读成一段漂移。**

签名三：它自我加固，越正规化越严重。这一条最要紧。证候标准越细致，条目定义越明确，医生在相邻两次之间做出相同判定的概率就越高——因为标准的作用正是压缩判定的随机波动。培训越严格、质控越强调前后一致、电子病历越是自动带出上一次的诊断，判同的倾向就越强。**规范化不是这个问题的解药，规范化是它的加速器。**一个手写的、每次重新描述的、彼此不一致的病历，吞差反而更少。

三条签名全部写满，这一格立得住。

六之二·第①格与第②格的分界只能由一次直接比较给出

第①格与第②格在整份病历上看起来完全相同：一串一样的证名，中间没有任何变更登记。区分它们只有一条路，而且这条路不需要任何新工具：

把首诊与最近一次的四诊记录抽出来，遮住日期与顺序，交给一位不知道来源的判读者，问他这是不是同一个证。

这个动作的成本极低，所需的材料在每一份病历里都已经存在。**它之所以从来没有被做过，不是因为难，是因为没有人要求过。**病历是按时间流水读的，前一次就在眼前，首诊在二十三页之前，而任何一次复诊的目的都不是回头看第一次。

七·一句不含情态词的问话

一个判断若要管用，必须能被压成一句问话，而这句问话必须能拿去问记录系统，不必拿去问人的判断。

本章的问话是：

把第一次记录与最近一次记录直接并排，不看中间，会不会判出同一个证？记录里登记过几次证型改变？

它不含任何情态词。它问的全是可以清点的东西。拿到五个不同的地方去问，答案互不相同。

问一份中医门诊的连续复诊档案：答案是——首末直接比常常判出不同的证，而中间登记的改证次数是零或一。**而这个比较从来没有人做过。**

问一份心理治疗的疗程记录：答案是——每次会谈记录多写“较前无明显变化”，而首末对比往往明显不同。值得注意的是，这个领域已经部分意识到了：临床总体印象量表明确要求评定“较基线的变化”，而不是较上次的变化。**这是一个反例，说明有人碰到过这个问题并给出了一个局部的解。**

问一份工业设备的状态监测日志：答案是——报警阈值基于与上次读数的差值，缓慢漂移不触发；而工程界为此专门发展了趋势分析与基线比对，把每一次

读数放回一条长期曲线上，并与出厂基线、与同型机队比。这里有账，而且账做得相当成熟。

问一份心理物理实验的原始记录：答案是——首末之间的最小可觉差数目，与序列比较中报告“不同”的次数，两者之差是这个学科的标准输出量之一。这里不但有账，这笔账就是这个学科的主要产品。

问一份儿童生长发育记录：答案是——每一次体检都可能写“较前无异常”，而百分位曲线显示持续下滑。生长曲线这个发明的全部意义，就是把每一次读数放回一条跨年的基线上，从而使相邻比较的累积失效显形。这是三个已解决案例里最有说服力的一个，因为它面对的对象与临床辨证最像：一个持续变化的活体，每次只能看一眼。

五个答案互不相同。其中至少两个不是从主张推出来的，而是查出来的：临床总体印象量表要求与基线比、生长曲线百分位追踪。本章在提出这条问话时并不知道这两处已经存在。

八·吞差率：一个跨领域量纲不同而比率相同的读数

有了问话，就可以有数。

先说清它要数什么。一段观察期有两端。从首端走到末端，若把中间全部遮住、只做直接比较，需要跨过若干个可分辨的档——记为 **k**。而这段期间的记录里，实际登记过的“改变”次数——记为 **m**。

$$\text{吞差率} = (k - m) / k$$

它是一个纯比值，没有量纲。它数的是：这段路上实际跨过的可分辨档数里，有多大比例没有被任何一次记录承认过。

在中医：k 由盲评者比较首诊与末诊的四诊记录得出——判出跨越了几个证型档；m 是病历中“证型改变”的登记次数。所需材料全部现成，只需要一次盲评。

在心理物理学：k 是首末刺激之间的最小可觉差数目；m 是逐次比较中报告“不同”的次数。这是这个学科的经典设计，数据成熟。

在工程状态监测：k 是首末两次状态之间实际跨越的可区分状态数；m 是诊断系统报出的状态变更次数。对应的现象叫状态漂移未被报警。

在**儿童生长监测**： k 是首末两次的百分位跨越的标准档数； m 是历次体检中记录“发育异常”或“需随访”的次数。

四处量纲完全不同——证型档、最小可觉差、可区分状态、生长百分位——而比率的含义完全相同。

八之一·它与既有指标结构性反向

一个新读数最容易犯的错误，是与某个既有主指标高度相关，于是被读成那个指标的代理。检验方法只有一个：举出一个“既有指标最好看、而这个数最难看”的真实例子。

本章能举出四个，而且四个的方向一致。

中医：一份二十四次复诊、证型始终记作脾气虚、评价者一致性极高、被当作辨证规范的模范档案——它的吞差率可以是一。 k 大于零而 m 等于零，比值为 1，这是理论上的最大值。反过来，一份被质控认为“辨证不稳定”、频繁改证的档案，若每次改证都对应一次真实跨档，它的吞差率是零。在**一致性**这个主指标上最难看的那份档案，可能是吞差最少的那一份。

心理物理学：一个被试的判断极其稳定、重测信度极高——这恰恰意味着他的阈限较大，因而在同一段刺激范围内他吞掉的差异更多。稳定与灵敏在这里结构性对立。

工程状态监测：一台设备的报警记录干干净净、可用率极高、维护记录漂亮——若它的报警逻辑只做相邻差值比较，这份漂亮记录与高吞差率完全相容，而且往往同现。

儿童生长监测：一个孩子每次体检都被判“发育正常”、体检记录无一异常，而百分位从第五十掉到第十。每一次都正常，全程掉了四十个百分位。

四个例子的方向一致：吞差率与既有主指标不但正交，而且在多数情形下**结构性反向**——既有指标越好看，这个数越大。这使它不可能被读成任何现有指标的代理。

八之二·它自带一件伦理

一个比率一旦进入制度，几乎必然被拿去考核人。因此必须把限制写死。本章写死三条。

第一，这个读数指的是记录结构，不是人。 吞差率高，说的是“这套记录方式吸收了多少”，不是“这位医生粗心”。同一位医生在采样间隔不同的两个病人身上，吞差率会差很多，因为它主要由采样间隔与分辨阈的比值决定，不由医生的态度决定。任何把吞差率直接换算为医师个人评价的做法都是误用。

第二，不得为了把这个数做好看而在安全关键的场合制造记录变动。 具体地说：不得以降低吞差率为由要求医生更频繁地更改证型判定。**更改证名是一个零成本的书面动作**，它可以在不做任何实质判定的情况下把 m 做大。这是本章预见到的主要滥用方式，第十六节会说清为什么它没有出口。

第三，已经按这个读数做出不利安排的，必须公开编码规则并给出复核通道。 吞差率的计算依赖两个判读：首末跨越几档，以及记录里哪些算“改变”。两个判读都不唯一。任何用它做出不利决定的机构，必须公开它用的判读规则，并允许被评价者提出另一种判读。

三条缺一，这个读数就不该被使用。

九·回到题目：何谓一个「证」

前面八节做的是地基。这一节回答题目。

九之一·证是链上的一段，不是集合里的一个类

分类学处理的是类。一个类由它的成员共有的性质定义，成员之间两两同类，类与类之间互斥。证候标准化的全部努力是把证做成这样的类。

本章的主张是：**做不成，而且做不成的原因不在技术。**

只要辨证是一个有分辨阈的判定系统——四诊的分辨力有限，这一点没有人否认——那么“分辨不开”这个关系就一定不传递。不传递的关系不能生成分割。**不能生成分割的东西，不是一个分类系统。**

那么证是什么？它是一条链上的一段。

这条链是这样构成的：把一个人可能处的全部身体状态排开，相邻状态之间在四诊这个分辨率下分辨不开，而距离足够远的状态分辨得开。这就是一条有序而相邻不可分辨的结构。**证不是这条链的自然关节，因为这条链没有自然关节。** 证是这条链被切开的地方——而切在哪里，由登记决定：某一次有人判了“变了”，那里就成了一个边界。

由此得到一句可以直接说出口的结论：

证的边界不是它自身的性质，而是这条链在何处被登记过一次改变。

这句话听起来像是把证说虚了。它不虚。它恰恰解释了几件长期解释不了的事。

九之二·三件长期解释不了的事，因此得到解释

第一件：证型为什么这么多，而且各家不同。若证是自然类别，各家分出来的类别应当收敛。事实是不收敛——不同的教材、不同的流派、不同的标准，切出来的证型数目与边界都不同，而且各自都能用。在本章的读法下这不是乱象：**链没有自然关节，切分点是约定的，因而不同的切法都可用，且不必收敛。**收敛的要求本身问错了问题。

第二件：为什么“证的转化”是中医的核心概念，而现代分类学里没有对应物。类别之间不发生转化——一个东西从一个类跳到另一个类，那是重新归类，不是转化。中医大量的核心内容是关于传变的：六经传变、卫气营血、脏腑相传。这些内容在把证当作类别的框架里，只能被处理成“归类的变更”，因而失去了全部内容。**在链的读法下，传变是这条链上的行走，它有方向、有速度、有路径，这些正是传变学说在讲的东西。**

第三件：为什么辨证一致性怎么做都上不去。若证是类别，一致性低说明标准不够明确；把标准做细，一致性应当上升。三十年的经验是上升有限，且到某个水平就停住。在本章的读法下这是必然的：**两位医生的分歧，大量不是“这不是脾气虚”的分歧，而是“这一次算不算已经越过了那个登记点”的分歧。**而登记点是约定的，把条目写细并不能消除关于约定的分歧。这条推论给出一个可检验的预测：**若把辨证任务改为“这一次与上一次比，跨过了几档”，一致性应当显著高于“这一次是什么证”的一致性。**前者问的是链上的距离，后者问的是切分点的位置。

九之三·“同一个证”这句话必须分成两句

中医语言里“同一个证”混用了两种关系，而这两种关系性质不同。本章认为这是两千年混乱的一个直接来源。

第一种是相邻同一：这一次与上一次，在四诊这个分辨率下分辨不开。这是一个**不传递**的关系。临床上绝大多数“还是这个证”的判断，说的是这一种。

第二种是跨段同一：两个时点，直接放在一起比较，判为同一个证。这是可以传递的——如果它是由直接比较得出的。

两者用同一个词，于是相邻同一被当成了跨段同一，一串相邻同一被成了一个跨段同一。**这一步在语言里是不留痕迹的，在记录里也不留痕迹。**病历上的第二十四“脾气虚”与第一个“脾气虚”，字面一样，而它们的来路完全不同：第一个是直接判定，第二十四是二十三次相邻同一的传递结果，而这个传递没有依据。

一条可执行的建议由此得出，而且它不需要接受本章的任何概念：**在记录中把这两种判断分开写。**相邻比较的结果记一栏，与基线比较的结果记另一栏。两栏都填，成本是每次多一次比较；而两栏不一致的那些记录，正是需要复核的那些。

九之四·与本系列同题另一章的分界

本系列另有一篇处理同一道题的论文，主张一个证是一次带处置授权的分层判定——它在同时成立的若干层里指定本次动手的那一层，其余为让层。

两篇的切面不同，而且是正交的两个切面：

那一篇切的是横断面：同一次判定内部，认定了几层、动了几层。本章切的是纵剖面：一串判定之间，跨了几档、登记了几次。

判决性的对照是这样的：拿一份“二十次复诊证型始终不变”的档案，那一篇要查的是每一次处置针对了几层，本章要查的是首末直接比较跨了几档。**两个数在数学上独立**——一个让层比为零的病例（每次只认定一层、也只动一层）完全可以有极高的吞差率，因为它每次都判“还是这一层”。反过来，一个让层比很高的病例也完全可以有零吞差率。

两篇合起来给出一个联合读法：**让层比高且吞差率高的档案，其证候疗效读数最不可信**——它既有看不见的搁置，又有看不见的漂移，而两者都会被计入同一个证候积分。这是两篇都无法单独给出的一条判断。

九之五·四个位置

把「证」与它周边的几个词收在一起：

位置	在本章的结构里是什么	在账上的形态
症	链上定位所用的坐标读数	四诊记录，已有完整字段

证	链上的一段，其边界为登记点	病历“证候”栏，字段有而边界的来路无记录
吞差	一串判同所吸收、跨次可加的差异量	无字段
传变	在链上的行走，有方向与速度	医案叙述中有，结构化记录中无
体质	链上一个长期停留的区段	体质量表，独立于病历，与链无接口

这张表里只有一行完全空白。而正是这一行的空缺，使“证型稳定”这四个字在临床记录中成为一个不可解读的读数——它既可能是第①格，也可能是第②格，而现行记录不提供任何区分二者的信息。

十·在别的地方，这个空缺长什么样

一个跨领域的判断，只有能在别的领域里据以预测一件具体的事，才算兑现。本节逐处兑现，共十四处。

一·心理物理学：吞差率在这里是现成的实验量。预测：同一批被试，采样步长越小（相邻刺激差越小），首末不一致而逐次全报“相同”的比例越高。这是一个可以在下午就跑完的实验，而且它检验的是本章最反直觉的那条推论——加密采样加重而非减轻吞差。

二·心理物理学：阈限越大的被试，吞差越多。预测：判断稳定性（重测信度）与吞差率正相关。这与“稳定即准确”的直觉相反，而这个相关在现有的心理物理数据集里应当可以直接算出来。

三·物理学：熵是“被判为相同的那些区别”的账，但它是一次粗粒化的账，不是一串判定的账。预测：任何试图把吞差直接写成熵的做法都会失败，因为熵不携带序。第十三节把这一处展开为对本章的削弱。

四·物理学：粗粒化的选择决定宏观态，因而“证型数目”是分辨率的函数而非病理的函数。预测：在四诊之外加入一组新的采集手段（例如舌象的定量成像），证型的数目应当上升而不是收敛——因为分辨率提高会把原本不可分辨的状态分开。若加入新手段后证型数目下降或收敛，本章这一处错。

五·工程学：可诊断性分析可以事先算出哪些状态对是不可区分的。预测：把这套分析移到四诊上——形式化地问“在四诊这套观测下，哪几组病理状态在原则上不可区分”——应当能得出一个有限的不可区分对清单。这个清单目前不存在，而它可以由现有的证候数据估计。有了它，就能事先知道哪些证之间的吞差风险最高。

六·工程学：状态监测的趋势分析是本章所要求的东西的成熟工程实现。它把每一次读数放回长期曲线、与出厂基线比、与同型机队比。预测：把这三条移植到中医复诊记录（与上次比、与首诊比、与同证型人群比），首末不一致而中间无变更登记的比例应当显著下降。这是一次可以在单个科室试点的改造。

七·工程学：滞回与去抖是为压制第④格而设的。这提醒本章：吞差率不是越低越好。工程界为减少误报警付出了大量设计成本，而滞回的代价正是提高吞差。第十三节展开为对本章价值排序的削弱。

八·儿科生长监测：百分位曲线是对抗相邻比较失效的成熟解。预测：在没有生长曲线的记录方式下（只记每次身高体重与“较前正常”），生长迟缓的识别应当系统性延迟，且延迟时长与体检间隔成反比而非正比——间隔越密，越容易被逐次吸收。这条预测可以在历史病历上回溯检验。

九·临床总体印象量表：明确要求“较基线”而非“较上次”。预测：在同一批病人上并行使用“较上次”与“较基线”两种评定，两者的分歧应当随疗程长度单调上升，而不随病情严重度上升。若分歧主要由严重度驱动，则它是评定难度问题，不是吞差。

十·精神科的诊断稳定性研究：一个诊断在随访中保持不变的比例被当作诊断效度的证据。本章给出相反的读法：诊断稳定性高也可能是吞差高。预测：在随访间隔较密的队列里，诊断稳定性应当高于间隔较疏的队列，而首末盲评的一致性应当低于后者。这两个方向相反的读数若同现，本章对。

十一·机器学习的概念漂移：模型在缓慢变化的数据分布上性能逐步下降而单次评估无异常。这个领域的对策正是本章所主张的——保留一个固定的基准集，定期与它比对，而不是只看滑动窗口内的指标。预测：只用滑动窗口监控的系统，其性能退化被发现的时点应当系统性晚于同时保留固定基准集的系统。

十二·财务与审计的重要性阈值：单笔低于阈值的差异不予调整，而多笔可累积。审计准则为此专门规定了未更正错报的汇总——低于单笔阈值的差异要被记入一张汇总表，期末合计后再判断是否重要。这是本章所要求的那个字段在另

一个领域里的完整实现，而且它有强制力。预测：把这个形态移植到临床记录（把每次“判同但有轻微变化”的观察记入一张汇总表，定期合计），复核触发率应当上升，而单次记录成本增加极小。第十七节的赌注照它写。

十三·法律上的持续性侵害与时效起算：每一次单独看都未达可诉门槛，累积后达到。法律体系对此发展了继续性侵权与时效的特别规则。分离线在于法律处理的是责任归属，本章处理的是记录结构；但两者共享同一个形式难题，而法律那边已经承认了累积的可诉性。

十四·语言学的音变链移：每一步音变都在相邻可辨范围内，若干代后整个元音系统移位而母语者从未察觉一次“变化”。这是本章的形式结构在人文领域里最完整的一个实例，而且它有一个关键补充：**链移的发现依赖于跨代的直接比对（历史文献或跨年龄组录音），而不依赖任何一代人的自我报告。**预测：任何依赖“当事人报告是否变化”的监测系统，在链移型变化上必然失效。

十四处里，第一、第八、第十二处给出的是可以立即执行、成本极低的检验；第三、第七处反过来限制了本章的主张。

十一·与最近的几种说法的分界

一个新提出的东西最大的风险不是被反驳，是被吸收。本节逐一处理最容易发生这种吸收的十种说法。每一种都给出一条可分辨的差别，以及一条能分出对错的观察。

堆垛悖论（古希腊，哲学）。把本章的主张剥掉全部中医名词，只留形式，它是这样的：一串各自通过的逐次判定，累积出一个整体上不成立的结论。这个纯形式的结构最古老的名字就是堆垛悖论——一粒沙不是堆，加一粒还不是堆，加到最后是堆。本章必须正面处理它，绕过去就是承认被覆盖。分离线在于它问**真值，本章问账**：堆垛悖论问的是“从哪一粒开始为真”，两千多年的处理方案（认识论的、超赋值的、多值的）全部围绕真值展开；本章不问真值，本章问的是**每一次判定吸收了多少、这个量记在哪里、它跨次可不可加**。堆垛悖论的任何一个解法都不给出这个量，因为它的框架里没有“记录”这个维度。若**观察到：为一份档案登记吞差量之后，关于“它是不是同一个证”的争论并未减少，则本章没有超出堆垛悖论，本章被它吸收**；若观察到争论从“是不是”转向“跨了几档、哪一次该登记”，则本章提供了一个它不提供的东西。

温水煮青蛙（通俗模板）。本章剥出的形式属于“每一步都可接受而整体不可接受”这一族，而这一族有一个人人都会说的通俗名字。必须点名并判定，否则读者读完就归到这里去了。分离线在于**这个模板讲的是主体没察觉，本章讲的是记录没承载**。温水煮青蛙的处方是“提高警觉”；本章的处方是“加一次直接比较”，而且本章明确预测提高警觉无效——因为二十三次判定没有一次是不警觉的，它们全都是对的。若观察到：**仅通过培训与提示（不改变记录结构）就能显著降低首末不一致率**，则这是警觉问题，本章错；若必须引入基线比对字段才有效，则本章对。这是一个可以设三臂的试验。

半序与不传递偏好（Luce 1956 等，数学心理学与决策论）。它说到这一步：当判断有阙限时，“无差异”关系不传递，需要用一种比等价关系更弱的结构来描述，并已给出完整的公理刻画与表示定理。这是本章形式基础的正式名字，本章承认自己站在它上面。分离线在于**它刻画的是关系的结构，本章提出的是一个可清点的量**：半序理论回答“这种关系长什么样”，不回答“一串这样的判定吸收了多少、怎么记”。若观察到：**从半序的表示定理可以直接导出本章的吞差率并给出记录方案**，则本章只是它的一个应用；若导不出记录这一层，则本章加了一样它没有的东西。这一条本章判自己只是站在它上面而非超出它太多——本章的形式层贡献接近于零，本章的贡献在账这一层。这句话必须写出来，否则是掩饰。

概念漂移（机器学习）。它说到这一步：数据分布随时间缓慢变化，模型性能逐步退化，而单次评估无异常；对策是保留固定基准集定期比对。分离线在于**它处理的是模型与世界的错配，本章处理的是判定序列内部的传递性失效**。两者的处方形式相同（保留基准），但概念漂移不要求那个量可加，也不关心“每一次判同吸收了多少”。若观察到：**把概念漂移的标准检测方法直接用于辨证记录就能给出吞差率**，则本章被它覆盖；若这些方法只能报“发生了漂移”而给不出跨了几档、哪几次该登记，则本章未被覆盖。

电子病历的复制前带与记录膨胀（医学信息学）。这是本章最危险的一位。它说到这一步：电子病历允许把上一次记录整体带出，医生只做少量修改，导致记录固化、内容失真、错误延续。已有大量实证文献。分离线在于**它是书写行为问题，本章是判定结构问题**——而这条分离线有一个干净的检验：**在严格禁止带出、要求每次独立重新采集与判定的系统里，吞差是否消失？**本章预测不消失，只是下降，因为相邻两个时点的状态确实分辨不开，独立重判也会判同。若观察

到禁止带出后首末不一致率趋近于零，则吞差就是复制前带，本章被吸收；若观察到显著下降而稳定在一个大于零的水平，则本章对，且那个残余水平正是吞差的真实大小。这是本章最重要的一条检验，因为它同时给出了吞差的测量方法。

模糊集与隶属度 (Zadeh 1965)。它说到这一步：边界地带的归属不是零一，而是一个介于其间的隶属度。分离线在于**它处理单次归属的程度，本章处理一串归属之间的关系**，而且隶属度不可加，吞差要求可加。更要紧的是模糊集不解决传递性问题——模糊等价关系仍要求传递性（以某种形式），而本章要的恰恰是放弃它。若观察到：**用模糊隶属度记录每次判定后，首末不一致而中间无变更的情形消失，则模糊化解决了问题，本章多余；若这些情形照样发生（只是每次的隶属度略有不同而无人合计），则本章对。**

中医的传变（《伤寒论》六经传变等）。这是中医内部与本章最近的说法，而且比本章早两千年。它说到这一步：证会转化为另一个证，转化有次第、有条件、有可预测的路径。分离线在于**传变已经假定了两端是可分辨的类别，并给出转变的规律；本章说的是在没有任何一次被记为“变”的情况下已经变了。传变有账——六经的次第本身就是账；吞差无账。两者不矛盾，而是互补：传变描述这条链的地形，吞差度量这条链上有多少路程没有被记录。若观察到：按传变规律预测的转归时点与病历上的证型变更登记时点系统性错开（登记晚于预测），则吞差存在，本章对；若两者吻合，则记录已经跟上了传变，本章的空缺主张为伪。这一条只需要一批有传变规律可依的病种档案即可检验。**

测量不变性（心理测量学）。它说到这一步：跨时点比较之前，必须先证明测量本身没变，否则比较无意义。这一位对本章既是近邻也是威胁。分离线在于**它是一个前置条件而非一个量：测量不变性检验的输出是“可比/不可比”，不是“吸收了多少”。但它的威胁是实质的——本章的首末直接比较，本身可能不满足测量不变性：二十四个月后医生变了、季节变了、病人的表述习惯变了。第十三节把这一处展开为对本章的削弱，并给出应对（盲评、剥离时间信息、多评者）。**

未更正错报的汇总（审计准则）。这是本章所要求的字段在另一个领域里的完整实现，而且它有强制力：单笔低于重要性阈值的差异不予调整，但必须记入一张汇总表，期末合计后重新判断。识别、认可、不调整、入表、期末合计、影响结论——六件齐全。**本章因此不能声称这个字段无人想到过。本章能声称的只**

是：在临床判定、心理评估、生长监测这几处，这个已经成熟的记账形态没有被移植过。这一位同时是本章的最好模板，第十七节的赌注照它的形态写。

一致性统计 (Kappa 等)。它说到这一步：把判同与判异计数，算出超出偶然的一致程度。本章与它的关系不是并列而是相反方向：**它把判同当作零信息，本章认为判同携带量。**分离线因此非常锋利：**同一份“二十四次判定完全一致”的档案，一致性统计判为最高质量，本章判为吞差率可能为一。两种读法对同一份材料给出相反的评价。**这不是“我比它多说了一点”，这是相反评价，因此分离线成立。若观察到：**一致性最高的临床单位其首末盲评不一致率也最高，则本章对；若两者同向（一致性高的单位首末也一致），则本章这一处错。**

十条之中，最近的一位是半序理论——本章的形式基础全部来自它，本章在形式层没有新东西。本章之所以仍不被它吸收，是因为它止步于关系的结构，而本章问的是一串这样的判定在记录上留下了什么、少了什么、少的那部分能不能加起来。这一步是形式理论不需要走的，因为它不面对记录。

十一之一·这些说法共同看不见的那一块

逐个问最近的几位“它把什么当作给定，因此结构性地看不见什么”，答案有一个共同的形状。

堆垛悖论把“要解决的是真值”当作给定，因此看不见被吸收的量是可累加的实在；一旦承认这个量，堆垛悖论就不再是一个语义问题，它的两千年文献要换问题。一致性统计把“判同不携带信息”当作给定，因此看不见一致性可以是吞差的产物；承认它，一致性统计作为质量指标的解释基础要重建。复制前带研究把“问题在书写行为”当作给定，因此看不见独立重判也会判同；承认它，这个领域的全部处方（禁止带出、强制修改、留痕）都只能解决一部分。传变学说把“两端是可分辨的类别”当作给定，因此看不见未被登记的行走；承认它，传变的次第就不再是一条可以直接对照病历核验的规律。

四条盲区背后是同一块地：

判定为相同，这个结果不携带任何量。

判同等于零，这是所有人共同站着的那一块。互相批评的各派都站在它上面，因此谁也回头看不见它。分类学在它上面，一致性统计在它上面，质控在它

上面，标准化在它上面。“未被任何学科辨识”的真正来源不是没人看，是看的人都站在它上面。

本章这一次只掀开了它的一角——只处理了同一分辨率下的一串判同，没有处理分辨率本身在这串判定中间发生变化的情形（医生换了、手段升级了）。这一角留在第十六节交代。

十二·与本系列各章的分界

本系列此前已有若干处理相邻问题的论文，其中一篇与本章使用了完全相同的三个学科，另一篇处理完全相同的题目。本系列篇目是彼此最近的读者，必须逐篇指名。

《阈前距》（《治未病立账学》第二章）（何谓未病，同样取材于物理、心理、工程这三处）。那一篇处理的是对象与判定线之间的距离，其两端只有一端被记账——对象在动，线也在动，而“这次变化里有多少来自线”不设字段；它的读数是线移占比。分离线在于**本章假定线不动**：本章全程假定判定标准不变、分辨阈不变，只问在同一把尺子下一串判同吸收了多少。同一份“连续二十年读数正常”的档案，那一篇要查的是判定标准的版本历史（这条线改过几次、哪一处、依据哪份文件），本章要查的是首末两端的直接比较（不看中间，还是不是同一个读数）。两条查法指向完全不同的档案。两者可以叠加，而且叠加时方向可能相反：一条向宽处移动的判定线，可以恰好抵消掉一段真实漂移，使首末直接比较也判为相同。这种情形下两篇都读不出问题，而两篇合起来可以——这是本系列两篇必须并读的一个具体场合。三家的立场在两篇里全部不同：那一篇的物理占当下序参量、心理占症状网络、工程占载荷与规定功能；本章的心理占共变结构与局部独立性、物理占粗粒化与尺度、工程占可诊断性与观测配置。

《让层》（第一章）（问题：何谓一个证，取材于教育、化学、艺术）。第九节之四已经给出分工：横断面对纵剖面，让层比对吞差率，两个数数学上独立，联合读法给出一条单篇给不出的判断。此处补一条判决性对照：一份让层比为零、吞差率为一的档案，那一篇会判“记录完整、无搁置”，本章会判“记录失真、漂移未登记”。两篇对同一份材料给出相反的评价。

《撑平》（《治未病立账学》第一章）（何谓未病，教育化学艺术）。那一篇说的是读数持平由持续双向抵消维持，被抵消的维持工作不设字段；读数是他的持率。分离线在于**对象动没动**：那一篇里对象被外力按在原处，本章里对象一路

在走而记录没跟上。同一份“症状长期不变”的档案，那一篇查是否有人在持续代偿，本章查首末是否还是同一个证。

《叠裕》（《治未病立账学》第三章）（余量按单位分账）。分离线在维度：那一篇是横向的（多个单位共用一份余量），本章是纵向的（一串判定共用一个“相同”的结论）。若一个现象在只有一个单位时仍然出现，它不属于那一篇；若一个现象在只有两次判定时仍然出现，它不属于本章——本章要求至少三次，因为传递性失效至少需要三个点。这条判据可以在任何具体案例上直接执行。

《功债》（《治未病立账学》第九章）《悬量》（《治未病立账学》第八章）《漏汰》（《治未病立账学》第六章）《欠返》（《治未病立账学》第五章）（治未病诸题）。这几篇处理的都是单条时间线上某一环节的读数结构。与本章的共同分界很简单：它们处理的是“一次动作之后留下了什么”，本章处理的是“一串判定之间少记了什么”。前者的对象是动作的残留，后者的对象是判定的吸收。

本系列这些篇目共用一个未言明的做法：都在找各领域账本上被当作垃圾、余数或噪声的那一格。这是它们共同的强处，也是共同的弱点——第十六节会交代这个弱点在本章身上的具体形状，而且它这一次特别难堪，因为本章自己的形式结构恰好可以用来诊断本系列自己。

十三·三处反过来削弱本章的地方

若三个来源全都顺着一条判断说，那它多半只是在这三处各挑了一句合用的话。本章有三处削弱，其中第一处动的是适用范围，第二处动的是价值排序。

第一处，物理学削掉了本章最想主张的那句话。本章原本会写下：被判为相同而实际存在的差异，在所有领域里都不设字段。物理学不允许这样写——熵就是这笔账，而且它是物理学最中心的量之一。玻尔兹曼数的正是“在这个粗粒化下，有多少微观区别被当作了同一件事”。这不是一个边缘概念，它是热力学第二定律的全部内容。

本章因此必须说清自己与它差在哪里，而这一说就削掉了很多东西。熵数的是一次粗粒化抹掉了多少，它是一个静态量，与顺序无关——同一个宏观态，不管系统是怎么到达的，熵一样。本章的吞差要求跨次可加且依赖顺序：二十三次判同吸收的量，必须能加起来，而且它们的排列顺序决定了首末的距离。

这个差别把本章的适用范围缩小了：**本章只适用于有时间序的、逐次比较的判定系统**。对于无序的分类系统——一次性把一批对象分成若干类，不涉及同一对象的重复判定——本章没有内容。而且物理学在这里给本章出了一道本章答不了的题：**吞差的可加性从哪里来？**本章的回答是“因为链是有序的”，这个回答成立，但它不能推出各次吞差的量可以简单相加——若各次的分辨阈本身随位置变化（心理物理学明确知道阈限随刺激强度变化），那么简单相加是错的。**本章在这一点上只能给出一个近似，并把它写进限制。**

第二处，工程学削掉了本章的价值排序。本章原本会写下：**吞差率应当被压低**。工程学不允许。状态监测领域为压制误报警投入了极大的设计成本——徘徊、去抖、多重确认、最小持续时间。这些机制的作用全部是**提高吞差**：它们故意让系统对小变化不响应，代价是漏掉缓慢漂移，收益是不被虚警淹没。工程界清楚地知道这是一个权衡，而且知道权衡点应当由误报与漏报的相对代价决定，不由某一方越低越好决定。

本章因此不主张吞差率越低越好。**在虚警代价高的场合，高吞差率是正确设计的后果。**这一处削弱直接改写了第八节之二的第二条伦理：不得以降低吞差率为由要求更频繁地更改判定，正是从这里来的。本章剩下的主张只有一句：**这个数应当被记下来，而它现在没有被记下来。**

第三处，心理学削掉了本章首末直接比较的地位。本章原本会把首末直接比较当作金标准——不看中间，直接比两端，得到 k 。心理测量学的测量不变性检验不允许这样。二十四个月之后，采集这份四诊记录的条件已经变了：可能是另一位医生，季节不同，病人对自己症状的表述方式也在这段时间里被前二十三次问诊塑造过。**在测量本身不可比的情况下，首末比较的差异有一部分来自测量而不是来自对象。**

本章接受这一处，并把它写成前置条件：**首末比较必须由不知道来源的评者、在剥离时间信息与顺序信息的条件下做，且需多评者。**这提高了 k 的获取成本，也削弱了本章原本“这个检验成本极低”的主张——**低成本的是把材料抽出来，不低成本的是把它评准。**本章在第十五节的低成本检验一栏因此做了调整。

三处削弱之后剩下的主张，比本章最初打算写的窄得多。这是应该的。

十四·不适用的边界

本章的判断在下列情形不适用，或需要重大修改：

无序的一次性分类。若不涉及同一对象的重复判定，就没有链，也就没有吞差。本章对横断面的分类学问题没有内容。

分辨阈远小于变化速率的场合。若相邻两次之间的变化明显大于分辨阈，每一次都会被判为“变了”，吞差趋于零。急性病、快速进展的病程属于这一类。本章的全部内容集中在慢性、缓变、长期随访的场合。

判定标准在观察期内发生变化的场合。本章全程假定尺子不变。尺子在动的情形属于本系列另一篇的地盘，两者叠加时可能互相掩盖，见第十二节。

分辨阈随位置显著变化的场合。第十三节第一处已经说明：此时各次吞差不能简单相加，本章的比率只是近似。

判定本身改变对象的场合。在心理治疗与某些教育场合，“判定为相同”这个动作本身会影响对象——告诉一个人“你和上次一样”，会改变他下一次的表現。此时链不是外在的，本章的结构需要重做。

只有两次判定的场合。传递性失效至少需要三个点。两次判定之间只有一个相邻比较，无所谓吞差。

十五·可以让本章翻车的观测

本章的主张可以被分层引用。**第一层是吞差这个构念**，最强也最易倒。**第二层是那张四格表**，即使构念被推翻，“相邻判同/首末可分辨”这个二维分类仍可能有用。**第三层是一条不依赖前两层的经验主张**：现行辨证一致性指标把判同当作零信息，因而它无法区分真稳定与漂移未登记。这一层最容易检验，且检验它不需要接受本章的任何概念。

十五之一·最强的竞争解释，以及排除它的条件

必须先写出那个最朴素、最难反驳的说法，不能挑一个好打的。它是：

其实不就是医生照抄上一次吗？

这个解释很强：它符合所有人的经验，有大量实证文献支持（电子病历的复制前带），而且处方很便宜——禁止带出、强制重新采集。

排除它的条件写成这个形状：

在严格禁止记录带出、要求每次独立采集与独立判定的条件下，若“首末盲评不一致而中间无变更登记”的发生率降至接近零，则本判断为伪；若该发生率显著下降而稳定在一个明显大于零的水平，且该残余水平与采样间隔呈剂量—反应关系（间隔越密、残余越高），则本机制成立。

关键在于两个处方不同，而且不同之处可以直接检验：

- 若只是照抄，**禁止带出**就应当解决问题。
- 本章预测禁止带出**不能解决**，且残余部分随采样间隔缩短而增大。**这条预测的方向与常识相反**——加密复诊在直觉上是更负责任的做法，而本章预测它加重吞差。这是本章最容易被证伪的一条，也是最值钱的一条。

十五之二·样本纪律

凡涉及剂量—反应关系的检验，样本至少六个单元，且在关键混淆变量上同质。**本章明确排除跨国或跨地区的两两比较作为证伪检验**——两地在制度、记录传统、医患关系上处处不同，所谓控制变量在方法上站不住，那样的比较只能作启发式说明。

可用的同质设计：同一家中医院的六个以上科室；同一省内多家同级中医院的同一专科；同一专科内按复诊间隔自然分层的多组病人（这一组尤其有用，因为自变量在同一制度环境内有真实变异）。

十五之三·其余可以让本章翻车的观测

一·若在辨证任务中把提问改为“较上次跨了几档”，其评价者一致性并不高于“这一次是什么证”的一致性，则第九节之二关于“分歧在切分点而非在链上位置”的主张为伪。

二·若一致性最高的临床单位其首末盲评不一致率并不更高（两者同向），则第八节之一的正交主张为伪，吞差率是一致性的一个代理。

三·若在四诊之外加入新的定量采集手段后，证型数目下降或收敛，则第十节第四处关于“分辨率提高会把不可分辨状态分开”的推论为伪。

四·若心理物理数据集中判断稳定性与吞差率不相关，则第十节第二处为伪。

五·若采用模糊隶属度记录每次判定之后，“首末不一致而中间无变更”的情形消失，则模糊化已解决问题，本章多余。

六·若按传变规律预测的转归时点与病历上的证型变更登记时点系统吻合，则记录已经跟上了传变，本章的空缺主张为伪。

七（正向读数，写成顺序而非相关）·若引入基线比对字段，那么“首末不一致被发现”的时点上升，应当早于“证型变更登记次数”的上升。发现在先，登记跟上在后。若观察到两者同时上升，或登记先动，则说明医生只是在多改证名而没有真的做基线比对——而这恰恰是第十六节预言的那种滥用，届时本章关于“缺的是比较而非意愿”的机制解释为伪。

八（低成本可实做，所需材料已经存在）·取任意一家中医院任意一批连续复诊满十二次的档案，抽出首诊与末诊的四诊记录，遮住日期与顺序，交多位盲评者判读。若绝大多数档案的首末被判为同一个证，则吞差现象不存在。这个检验不需要采集任何新数据；需要成本的只是盲评的组织，而非数据的获取。

若本章被证伪，学到的东西是明确的：那意味着临床辨证的相邻判定具有足够的传递性，因而把证做成互斥类别的方向是对的，只是精度不够。这会把资源引向与本章完全相反的地方——继续细化标准，而不是引入基线比对。

十五之四·本章解释不了的两个洞

第一个洞：吞差在什么条件下会突然显形？临床上“病情突然转变”的那些时刻，多少是真的突变、多少是累积到无法吸收？本章说这个误读的方向是固定的，但给不出比例，也给不出显形的阈值。

第二个洞：当分辨手段本身在观察期内升级时（新的检测进入四诊之外），过去的那些判同要不要重算？本章没有处理这种情形，而它在现实中正在发生。

十六·本章自己的吞差率

按本章自己的判据给本章定位，结果很难堪，而且必须说出后果。

本系列这一系列篇目，每一篇与上一篇之间都做过一次相邻比较，而每一次的结论都是“这是新的一篇”——三个不同的学科、一个不同的题目、一个不同的比率、一个不同的赌注。相邻比较全部通过。

而首末直接比较从来没有做过。把本系列最早的一篇与本章抽出来，遮住题目与学科，交给一位不知道来源的读者，问他这是不是同一篇文章的两个版本——本章无法预先声称他会答“不是”。

这个诊断本章自己躲不掉，理由有三个，逐一说清。

后果一：本章正落在自己指认的那一格里。第六节说那一格的三条签名是短期指标改善、失效不产生信号、越正规化越严重。本系列完全符合：每一篇单独看都锋利（短期好看）；每一次“这是新的一篇”都没有错，因而没有任何一次可以被追责（无信号）；而本系列的方法越成熟、写法越固定、格式越统一，相邻两篇判同的可能性就越大（越正规化越严重）。本章这一次做的事，恰好是它自己描述的那个结构的又一个实例。

后果二：本章的吞差率无法由作者自己算。k 需要盲评，而作者不能盲评自己。本章因此不能声称自己不在第②格里，只能声称这个检验应当被做，且本章接受它的结果。这是本章能给出的最诚实的位置——它比“本章也有局限”这类话要具体，因为它指定了一个具体的、可以让本章难堪的动作。

后果三：本章对本系列的建议必须先用在在本系列自己身上。本章第九节之三建议在记录中把相邻比较与基线比较分开写。移到本系列就是：每一篇不但要与上一篇划界，还要与本系列第一篇直接划界。本章第十二节做了前者，做后者只做了一部分。这是一处没有做完的工作，本章承认它。

十六之一·反噬预言，且没有出口

吞差率一旦进入考核，最省事的达标法不是做基线比对，而是多改几次证名——把 m 做大， $(k-m)/k$ 就小了。

而改证名是一个零成本的书面动作。它不需要重新采集，不需要重新判定，不需要任何人同意。一位医生完全可以在不改变任何实质判断的情况下，把“脾气虚”轮流写成“脾虚”与“脾虚湿困”，使记录上的变更次数上升。

结果会是双重的损失：吞差率的数字变好看，而“证型稳定”这个本来携带临床意义的信息被彻底毁掉——因为它再也不能被读作病情稳定。

有没有出口？本章找过三条，三条都不通。

第一条是用盲评核验 m 的真实性——但 m 是逐次的，核验每一次改证的实质性，成本高于整个测量本身。

第二条是只用 k 不用 m，即只考核“首末是否一致”——但那样会惩罚真实的病情变化，把第③格与第②格混为一谈。

第三条是不把它用于考核而只用于研究——但一个数一旦存在于系统里，它进入考核只是时间问题。

根本的困难在于这个比率的分子由医生自报（m 是记录里的登记次数），而分母需要外部盲评（k 需要独立判读）。任何分子自报、分母外评的比率，都会从分子上被攻破。我看不到出口。本章能做的只有第八节之二那三条限制，而三条限制全部依赖执行者的自我约束——这恰恰是本章自己论证过的最不可靠的东西。

十七·一条写死日期的赌注

一个判断若不肯把自己压在一个可以到期兑现的事实上，它就还只是一种说法。本章压这一条：

在二〇三六年十二月三十一日之前，中医电子病历的国家级或学会级结构化标准（或中医病历书写规范的修订版）中，将出现一个独立于“较前变化”的基线比对字段——要求复诊时的证型判定同时与上一次和一个固定基线（首诊或指定基线次）比对，并分别登记两个结果；且“基线比对判为不同而中间无变更登记”的记录被定义为一类需要复核的记录。

赌注若不写死什么不算命中，任何模糊的进展都可以被宣布为兑现。以下五条写死：

一·只要求“参考既往病历”或“结合病史”而没有独立的基线比对字段与结果登记的，不算命中。参考既往病历现在就已经是要求。

二·仅在科研病例报告表、课题专用病历或单一机构内部系统中出现，未进入常规诊疗规范的，不算命中。

三·基线比对由同一位医生自评、且没有盲法或独立复核安排的，不算命中。这一条最要紧。第十六节已经论证：这个比率的分子自报、分母外评，突破口就在自评。一个由本人自评的基线比对字段，等于把 k 也交给了 m 的那只手。

四·只登记“较基线有变化／无变化”而不登记跨越档数的，不算命中。二值登记无法给出 k，因而无法算出这个比率；而不能算出比率，这个字段就只是又一个提醒。

五·把“证型稳定率”设为正向考核指标的，不算命中，而且构成对本章主张的否定。证型稳定率正是本章指认的加重因素——它奖励的恰好是判同。若最终落地的制度一边加了基线比对字段、一边继续以证型稳定率考核，那么这个字段会被用来证明稳定是真的，而不是用来发现漂移。

这条赌注的形态照着审计准则里“未更正错报的汇总”写的：识别、认可、不调整、入表、期末合计、影响结论。那一处六件齐全且有强制力。本章赌的是这套形态会被移植到临床判定上。移植失败也是一个结论——它会说明临床判定与财务报表在结构上有一处本章没有看见的根本不同，而最可能的那一处是：**报表有一个明确的期末，临床没有。**

十八·结论

本章从一道两千年的定义题出发，但没有去争边界该划在哪里，而是退一步问：“**同一个证**”是不是一个可传递的关系？

三个互不通约的传统——心理测量学的共变结构判据、物理学的粗粒化判据、工程学的观测配置判据——在“该由什么裁定”上势不两立，而在一件事上完全一致：**给定分辨标准之后，判定为相同的那些可以被归进同一个类。**也就是说，它们共同假定这个关系可传递。

推翻它的材料就在其中一家自己手里。心理物理学用一百五十年证明了：任何有阈限的判定系统，它给出的“分辨不开”关系都不传递——相邻两两不可分辨的序列，两端明确可分辨。而临床辨证正是一个有阈限的判定系统。

由此得到的主张是：一个「证」是一条由相邻不可分辨关系串起来的链上的一段，它的边界不是它自身的性质，而是这条链在何处被登记过一次改变。与之相伴的是一类在三套账上都不设字段的东西——吞差：在一串各自合法的判同里被逐次吸收、在任何一次记录中都不出现、而跨次可加的那些差异量。

两根轴给出四个格：相邻判同而首末不可分辨的是真稳定，相邻有判异而首末不可分辨的是正常追踪，相邻判异而首末不可分辨的是过度反应——三格都有制度位置。只有相邻判同而首末可分辨的那一格，既有后果又无位置，而且它通过全部形式审查：短期指标上表现为改善，失效不产生任何信号，且越规范化越严重。

一句不含情态词的问话把它落地：**把第一次记录与最近一次记录直接并排，不看中间，会不会判出同一个证？记录里登记过几次证型改变？**由此得到一个跨领域量纲不同而比率相同的读数——吞差率，等于未被任何记录承认的跨档比例。它与辨证一致性结构性反向：一份二十四次证型始终一致的模范档案，吞差率可以是一。

回到题目：证不是集合里的一个类，是链上的一段；证型不收敛不是乱象，是链没有自然关节的必然后果；传变不是归类的变更，是在链上的行走；而辨证一致性上不去，是因为分歧大量落在“这一次算不算已经越过登记点”，而登记点是约定的。中医语言里“同一个证”混用了相邻同一与跨段同一两种关系，一串相邻同一被当成了一个跨段同一，而这一步在语言与记录里都不留痕迹。

三处削弱把主张收窄了：物理学的熵证明“被判为相同的差异”并非无人建账，本章只在跨次可加这一形态上是新的，适用范围因而缩到有时间序的逐次判定系统；工程学的滞回与去抖证明高吞差率有时正是正确设计，本章因此不主张压低它，只主张记下它；心理测量学的测量不变性削掉了首末直接比较的金标准地位，本章因而必须要求盲评与剥离时间信息，检验成本随之上升。

本章与本系列另一篇同题论文的关系是正交的：那一篇切横断面（一次判定内部认定了几层、动了几层），本章切纵剖面（一串判定之间跨了几档、登记了几次）。两个数数学上独立，而联合起来给出一条单篇给不出的判断——**让层比与吞差率同时高的档案，其证候疗效读数最不可信。**

本章自己的吞差率无法由作者自己算，因为分母需要盲评。本章因此不能声称自己不在那一格里，只能声称这个检验应当被做。而本章的处方一旦进入考核，最省事的达标法是多改几次证名——分子自报、分母外评的比率，总是从分子上被攻破。**对此我看不到出口。**

最后是那条赌注，压在二〇三六年十二月三十一日：一个独立于“较前变化”的基线比对字段，要求同时登记与上次、与基线的两个比对结果，并把“基线判为不同而中间无变更登记”定义为需要复核的记录。五条不算命中已经写死，其中最要紧的一条是：**由同一位医生自评而无独立复核的基线比对，等于把分母也交给了分子那只手。**

第三章 配格

——一个「证」为什么只做到“同格”而没有做到“可互换”：论“同证”与“同方”之间那一步在三套账上为何不设字段

提要 同一个证，同一张方，在两个病人身上疗效不齐——这是中医临床的常态，也是它在现代评价体系里最难过的一关。通行的解释是辨证不够精细，因而处方是继续细分：分兼夹、分体质、分地域、分年龄。本章认为这条路走不通，而且走不通的原因不在精度。本章提出：“被归为同一类”与“在使用处可以互相替换”是两件事，而中医的记录只承载了前者。一个「证」不是由发生史裁定的类型，不是落在规格带内即可互换的规格，也不是由共享计量制度维系的可比类别，而是一次只做到“同格”而未做到“可互换”的判定——它把个体归入一个格，而决定这一张方与这一个人配不配得上的，是这个格内部的位置。这个位置本章称为**配格**。推翻上述共同假定的材料来自工程学自己：精密装配里的分组选配与配对件证明，两个都落在同一公差带内、检验全部合格的零件，可以不能互换。由此得到一个跨领域量纲不同而比率相同的读数——**配格率**。它与指南依从性结构性反向：一份完全按标准方、零加减的处方，依从性满分而配格率为零。本章并指出，“辨证论治”这四个字实际包含两个动作——分类与配对，而“方证相应”这个说法把它们压成了一个；输血医学早已把这两步分开（血型分类与交叉配血是两个独立步骤、两份独立记录），中医尚未。

一·一个每天发生而无人当作问题的现象

两位病人，同一天，同一位医生，都辨为脾气虚。医生给第一位开四君子汤加黄芪、升麻，给第二位开四君子汤加陈皮、半夏。两张方的底子一样，加减不同。

问医生为什么不同，他会说：第一位气陷明显些，第二位湿重一点。

再问：那他们是不是同一个证？他会说：是，都是脾气虚。

这里有一件事从来没有被当作问题：如果他们是同一个证，为什么方不能一样？如果方不能一样，“同一个证”这四个字究竟保证了什么？

通行的回答有两种。第一种说这是个体化，是中医的长处，不必深究。第二种说这说明辨证还不够细——真正的证应当包含“兼气陷”与“兼湿”，把证分得更细，同证就能同方了。

第二种回答是过去三十年证候标准化的主流方向，而且它非常合理：分类的目的本来就是让同类的成员可以被同样地对待。若同类成员不能被同样对待，那说明类没有分好。

本章要论证的是：**这条路走不通，而且走不通的原因不在精度。**

理由可以先用一句话说清：**把类分得足够细以保证同类可同方，其终点是每个人一个类；而每个人一个类，就不再是分类。**这不是一个滑坡论证，它是一个可以直接检查的结构：只要个体之间的差异是连续的、多维的，任何有限的分类都必然在类内留下差异，而这些差异是否影响处置，与类分得多细无关——它取决于处置手段对这些差异有多敏感。

中药复方对个体差异极为敏感。这不是缺点，这是它的作用方式决定的。一张方是十几味药的配伍，任何一味的增减都改变整体的偏性；而人的偏性是连续分布的。**在这样一对手段与对象之间，分类永远不可能细到保证互换。**

于是问题就变了。它不再是“证该怎么分才对”，而是：

“归为同一个证”这个判定，究竟保证了什么，又没有保证什么？没有保证的那部分，记在哪里？

答案是：不记在任何地方。而且它不是记得少，是不算一样东西。

本章的路线：第二节请三个互不通约的传统各出一条最强的答法；第三节找出它们底下那个谁都没说出口的共同假定；第四节从其中一个传统自己的材料里找出推翻它的那一件；第五节到第八节给出主张、两根轴、一句可以拿去问记录系统的问话，以及一个可以在几个领域并排比较的数；第九节回到题目；其余各节交代兑现、与相近说法的分界、削弱、边界、可以让本章翻车的观测，以及一条写死日期的赌注。

二·三条不肯让步的答法

“两个东西凭什么算作同一类”，有三个传统给出过强硬且互不相容的答案。本章取它们最强的版本。

二之一·第一条：由它是怎么长出来的裁定

这一条来自细胞学，而且它正处在这个领域当下最激烈的争论中心。

单细胞测序普及之后，“细胞类型”这个词出了大问题。原来靠形态与几个标记物划出来的类型，在高维表达谱上变成了连续的云——中间态到处都是，同一个“类型”内部的异质性常常大于类型之间的差异。于是有人主张放弃类型，只谈状态。

而这个领域最有分量的一支回应不是这样。它主张：**细胞类型是实在的，但它的同一性判据不在表达谱上，在发生史上。**两个细胞属于同一类型，当且仅当它们的身份由同一套在进化中产生的核心调控装置维持。表达谱会随环境波动——那是状态；调控核心在个体发育与物种演化中被继承——那是类型。

这条主张的锋利之处在于它敢于接受一个反直觉的推论：**两个表达谱几乎完全相同的细胞，若它们的调控核心不同源，它们不是同一类型。**趋同现象在生物学里比比皆是，看起来一样不等于是一样。反过来，两个形态与功能差异很大的细胞，若共享同一个调控核心，它们仍是同一类型的两个状态。

把这一条移到中医上，它对应的是这样一种读法：脾气虚是不是一个真类型，取决于这些人的状态是不是由同一个病机发生过程产生的。**症状雷同不算数，来路相同才算数。**

它的主张压成一句：**看它是怎么长出来的就够了。**

二之二·第二条：由它当下的量测值落在哪一带裁定

这一条来自工程学，而且它是现代工业的地基，不是一个边缘意见。

十九世纪之前，机器的每一个零件都是配着做的——这一根轴配这一个孔，换一个就装不上。互换性制造改变了这一切：**规定一个公差带，凡是落在带内的零件一律视为同一规格，任何一个都可以装在任何一台机器上。**备件体系、大规模生产、维修产业，全部建立在这一条上。

工程学在这里的立场非常干脆，而且它明确排除了另外两条路：

它排除制造史。这个零件是哪个厂做的、用什么工艺、经了几道工序，与它是不是这个规格无关。图纸上不写工艺，只写尺寸与公差。**这正是互换性的全部意义——若还要追问来路，就回到了配着做的时代。**

它排除制度约定。装不装得上不是商量出来的，是可以当场试的。一个零件是不是合格，用量具一测就知道；两个零件能不能配，装上去一试就知道。工程学把这看作物理事实，不看作协调结果。

把这一条移到中医上，它对应的正是证候标准化那条路最纯粹的形态：制定条目、设定阈值、凡是达标的一律归为该证，至于这个人是怎么走到今天的、是哪位医生看的，一概不问。

它的主张压成一句：看它当下的量测值落在哪一带就够了。

二之三·第三条：由它嵌在哪一套共享的计量制度里裁定

这一条来自管理学。

管理学处理过一个工程学不曾正面处理的问题：当两个单位报上来同一个指标时，这两个数是不是在说同一件事？答案往往是否定的，而原因不在数据质量。两个分部的“销售额”可能在收入确认时点、退货处理、内部交易抵消上采用不同口径；两家医院的“平均住院日”可能在转科、出院登记、日间手术的计入上不同。没有一套共享的操作定义与计量规则，同一个名字在两个地方不指同一样东西。

管理学由此得出一条相当冒犯常识的结论：类别的可比性不是被发现的，是被建造的。它需要一套制度：共同的定义、共同的采集规程、共同的修约规则、共同的审核。这套制度撤掉，可比性立刻消失，而对象本身没有任何变化。

把这一条移到中医上，它对应的是这样一种读法：脾气虚在两位医生口中是不是同一个证，取决于他们是不是共享同一套四诊操作规程、同一套条目定义、同一套判读约定。不共享，这四个字就只是同一个词，不是同一个东西。

它的主张压成一句：看它此刻嵌在哪一套共享的计量制度里就够了。

二之四·三条为什么不能同时成立

第一条与第二条正面互斥。细胞学说同一性由发生史裁定，表达谱相同而来路不同就不是同一类；工程学说制造史完全无关，只看当下的量测值。若前者对，那么任何一套只看当下条目的证候标准在原则上都判不出真类型；若后者对，那么追问来路就是回到工业化之前。

第二条与第三条构成根基与病灶的关系，这是最值钱的一对。工程学每天在做的那件事——**照图纸检验**，把合格与否当作可以当场验证的物理事实——正是管理学诊断为问题所在的那件事：在管理学看来，“照图纸检验”这个动作本身依赖一整套共享制度（量具的溯源、测量方法的约定、修约规则、争议时以谁的仪

器为准），把它当作纯物理事实，正是把制度约定误认成了自然。两边不是意见不同，一边的地基是另一边的病灶。

第一条与第三条互斥。细胞学认为存在一个自然的裁决者——进化把调控核心的同源关系固定下来了，这不需要任何人同意；管理学认为没有自然裁决者，只有协调需要，类别的边界由使用它的人们为了合作而共同维持。若前者对，管理学关于“可比性是被建造的”就只适用于人造类别；若后者对，细胞学的“同源”本身也是由一套共享的判读规程产出的。

三条不能靠判谁对来化解，因为它们不在同一个语域里作答。要处理它们，只能往下挖。

三·三条答法共同站着的那块地

三条答法争的是同一件事：“这两个东西算不算同一类”该由什么裁定。一条说由发生史裁，一条说由当下量测值裁，一条说由共享制度裁。

它们争得很凶，而有一样东西三边谁也没有提出来过，因为在它们看来这根本不需要提：

归为同一类，就意味着在需要它的地方可以互相替换。

也就是说，分类的意义在于可替换性；一旦裁定两个东西同类，关于“能不能拿这个代替那个”的问题就已经被回答了。

这句话念给三边听，三边都会说这还用说吗。

工程学那一边最明白：互换性就是分类的目的。规定公差带的全部理由，就是让带内的零件可以互相替换。这一条不是工程学的推论，它是工程学制定分类的动机本身。

细胞学那一边：同一类型的细胞被认为在功能上可以互相顶替，这是再生医学与细胞治疗的前提——若同类型的细胞不能互相替换，“把这一类细胞补进去”这个整个治疗思路就不成立。

管理学那一边：标准化的目的就是让同一岗位、同一流程可以由不同的人执行，让同一指标可以在不同单位之间比较与合并。可比即可合并，可合并即在统计与决策中可互相顶替。

三边在“该由谁裁”上势不两立，在“裁完之后就可以互相替换”这一点上完全一致。而中医的证候研究整个建立在同一块地上：证候标准的目的是让同证的病人可以被同样地治疗，随机对照试验按证型分层的前提是同层内部可以合并计算，中成药与协定处方的存在依据是同证可以同方。**这些做法全部预设了归类蕴含可替换。**

这块地是本章要挖开的那一块。挖它不能从外面搬一个理由——那样只是加入争论，成为第四种立场。要挖开它，必须用三边之一自己已经掌握、已经发表、只是从来不住这个方向用的材料。

四·推翻它的那件材料，来自工程学自己

工程学有一套工艺，任何一本装配工艺手册都写，任何一家精密制造企业每天都在做，而没有人把它当作一个关于“分类”的命题。

分组选配。

情形是这样的：一对配合件——比如活塞与缸套、滚子与滚道、阀芯与阀体——设计上要求配合间隙极小，比如零点零零二毫米。而以当前的加工能力，单件的尺寸公差只能做到零点零一毫米。如果按互换性原则装配，把任意一个活塞装进任意一个缸套，间隙的分布会从零到零点零二毫米，远超要求。

工程学的解决办法不是提高加工精度——那太贵了。办法是：**把合格的零件按实测尺寸分成若干组，打上组号，只允许同组的件配在一起。**

现在请注意这套工艺在做的事情究竟是什么：

每一个零件都合格。检验全部通过。它们全都落在同一个公差带内，按图纸它们是同一个规格。而它们不能互换。

更极端的形态是配对件。某些精密偶件——高压油泵的柱塞与柱塞套、某些光学镜组——是成对研磨出来的，装配之后打上同一个编号，**一旦拆散就必须报废，即使两件各自都完全合格。**手册上会明确写“成对更换，不得单换”。

这件材料在工程学自己的领域里回答的是另一个问题——它回答的是“加工精度不够时怎么保证配合精度”，是工艺问题。**从来没有人把它拿来当作关于“归类是否蕴含可替换”的证据。**

但它就是那件证据，而且它硬到没有讨论余地：在工程学里，“合格”与“可互换”早已分家，而工程学自己为此付出真实的成本——分组、打号、分箱存放、成对管理，全都是钱。

而最要紧的一点在这里：组号是被记录下来的。它打在零件上，写在装配卡上，进入备件管理系统。工程学不但知道“合格不等于可互换”，它还为此造了一个字段，而且这个字段是强制的——没有组号，这个零件就不能被装配。

推翻之后，问题就变了。原来的问题是“证该由哪一维裁定”，现在的问题是：

如果归为同类不蕴含可以互换，那么决定“这一个能不能替换那一个”的那部分信息，去了哪里？

五·主张：配格

本章的主张是一句三重否定加一个新东西：

一个「证」不是由发生史裁定的类型，不是落在规格带内即可互换的规格，也不是由共享计量制度维系的可比类别，而是一次只做到“同格”而未做到“可互换”的判定——它把个体归入一个格，而决定这一张方与这一个人配不配得上的，是这个格内部的位置。

这个位置，本章称为配格。

三个否定各对应一边。第一个否定的是发生史判据，第二个否定的是规格带判据，第三个否定的是共享制度判据。三边都不会同意这条主张——三边都会说，你说的只是分类还不够细。但三边的证据合起来只能得出它。

五之一·配格不是什么

配格不是兼夹证。兼夹是在证名上加一个证名——脾气虚兼湿困。它做的是换一个更细的格，而不是记下格内的位置。加一层格，格内仍然有位置；再加一层，还是有。兼夹是沿着“分得更细”这条路走，配格是承认这条路有终点而问题在终点之外。这是两个方向。

配格不是体质。体质是跨时间稳定的个体倾向，它本身也是一套分类——九种体质仍然是九个格。把体质加进来，只是把格切得更细一点，格内仍有位置。

而且体质与配格的关系是可以直接检验的：若加入体质分型之后，同证同体质的两个人可以互换处方，则配格被体质吸收，本章错。第十五节把它列为可以让本章翻车的观测之一。

配格不是个体差异。“个体差异”是一个把问题命名之后就不再追问的词，它不指向任何可清点的东西。配格是可清点的：**它就是这一次为了让方配得上人而做的调整，在处方上是加减，在记录里有实物。**第八节把它变成一个数。

配格不是辨证不准。这一条最要紧。若两位同证病人的方不能互换是因为辨证不准，那么提高辨证水平就应当使加减减少。本章预测**不会**：一位辨证水平极高的老中医，他的加减不但不少，往往更多、更精细。**加减多不是辨证不准的症状，是配对做得细的症状。**这条预测与通行看法方向相反，第十五节列为可证伪点。

配格不是随意发挥。加减不是没有依据的。医生说得出他为什么加黄芪、为什么加陈皮。问题不在于有没有依据，而在于**这个依据不进入任何结构化字段**——它进入的是这一张方本身，而方是一个结果，不是一份关于配对依据的记录。

五之二·它为什么是一类原来不存在的东西

有一个检验很硬：删掉这个读数之后，它还在不在？

病历这一份账：处方栏里有一张方，辨证栏里有一个证名。加减的药味躺在方里，任何人都看得见。但**没有任何字段记录“这几味是为了什么加的、针对这个人的哪一个特征”**。方是结果，从结果反推依据，只有开方的人自己做得到，而且他很快也会忘。

证候标准这一份账：证候诊断标准的输出是一个证名或一组证名。它的结构里没有“格内位置”这个位置——它是一套归属规则，归属规则的输出是归属，不是位置。**要它输出位置，等于要它变成另一样东西。**

疗效评价这一份账：这一份账的漏洞最严重。随机对照试验按证型分层，同一层内部的病例被合并计算。**而同一层内部的病例，各自的方是不同的——加减不同。**合并计算的前提是同层内可互换，而这个前提从未被检验。更常见的做法是干脆固定处方、禁止加减，以保证“干预一致”，**这等于把配格信息从设计上删掉，然后惊讶于疗效不如临床观察。**

三份账都不设字段。所以在这三份账里，配格不是记录得少，是不算一样东西。删掉本章这个读数，它不会变成难以测量的东西，它会回到它现在所在的地方——藏在一张方的加减里，无人清点，无人传承。

五之三·一个立刻可以做的推论

如果配格是真的，有一件事应当已经在发生：

同证病例的处方之间，其加减部分的重合度应当显著低于其基础方部分的重合度；而这个差距不随医生水平提高而缩小，反而随水平提高而扩大。

这个预测不需要新数据。任何一批带处方的门诊病历，按证型分组，把每张方拆成基础方与加减两部分，分别算重合度，两个数一比就有答案。这个切法从来没有人用过，因为处方在研究里一直被当作一个整体。

六·两根轴与四个格

第一根轴是两个病例是否被判为同一个证。第二根轴是两张方能否互换。

	方可互换	方不可互换
判为同证	① 真同类——分类完全兑现，同证可同方	② 配格——同证而方不可换，差别不在证名里
判为异证	④ 分类过细——两个证名而同一张方通用	③ 正常——分类正确反映了处置差异

四格都不是虚设。

第①格是分类的理想状态，也是中成药、协定处方、临床路径成立的条件。它确实存在——某些证型、某些病程阶段，同证确实可以同方。

第③格是分类正常工作的样子：证不同，方不同，证名承担了它该承担的信息。

第④格是一眼可见的冗余：两个证名而处方相同，说明这两个格不必分。异病同治的某些情形接近这一格，而它不需要任何新框架就能看见——任何人比対一下处方就发现了。

第②格是本章要打的那一格。

六之一·为什么要打的是第②格

选定要打的那一格有一条判据：它必须是全部形式审查都能通过的那一格。一眼可见的坏不需要新框架。

第②格通过全部形式审查。辨证有据，方证相符，理法方药一线贯通，病历完整，加减有理。翻遍现行的病历质控条目与教学评价标准，没有一条会因为第②格而扣分——因为质控只查这一张方对不对得上这一个证，从不查这一张方能不能给同证的另一个人。

它还满足另外三条签名。

签名一：它在短期指标上表现为改善。个体化加减确实提高疗效，这是临床共识。而疗效好会被记在证名头上——“辨证准确，效如桴鼓”。证名因此承担了本不属于它的功劳：真正起作用的一部分是配格，而账记在了分类上。这个错记的方向是固定的：配格做得越好，证候辨识体系看起来越有效。

签名二：它的失效不产生任何信号。当一张方在同证的另一个人身上无效，会发生什么？它会被归因为“辨证不准”、“病情有变”、“个体差异大”。它不会被归因为“同证本来就不保证同方”，因为这个可能性在现有的解释框架里不存在。于是每一次失效都被读成分类没做好，处方是把分类做得更细——而这恰恰是本章认为走不通的那条路。误读的方向是固定的：总是被读成分类问题，从不被读成配对问题。

签名三：它自我加固，越正规化越严重。这一条最要紧。证候标准越细致，医生越相信同证可以同方；临床路径越规范，加减的空间越小；中成药与协定处方越普及，配格信息被删除得越彻底。标准化在这里不是中性的记录方式，它的目标就是消除加减——因为加减是可比性的敌人。而消除加减，等于把配格从系统里删掉，同时保留“同证同方”这个未经检验的假定。规范化不是这个问题的解药，规范化是它的执行者。

三条签名全部写满，这一格立得住。

六之二·第①格与第②格的分界只能由一次交换试验给出

第①格与第②格在一份病历上看起来完全相同：一个证名，一张方，理法通顺。区分它们只有一条路，而且这条路不需要任何新工具：

把这张方原样拿给同一时期、同一位医生辨出同一个证的另一位病人，问他会不会改，改几味。

这个动作的成本极低，材料在每一间诊室里都现成。它之所以从来没有被系统地做过，不是因为难，是因为没有人要求过。每一次诊疗都是围绕一个病人展开的，没有任何环节会把另一个病人的方拿过来问一句“这张能不能用”。

值得注意的是，医学内部已经有一个领域把这一步建制化了，而且是被血的教训逼出来的。第十一节会正面处理它。

七·一句不含情态词的问话

一个判断若要管用，必须能被压成一句问话，而这句问话必须能拿去问记录系统，不必拿去问人的判断。

本章的问话是：

这一张方里，有几味在基础方之外？这几味针对的是这个人的哪一个特征，记在哪个字段？同一时期辨出同一个证的另一位病人，能不能用这一张方？

它不含任何情态词。它问的全是可以清点的东西。拿到五个不同的地方去问，答案互不相同。

问一份中医门诊病历：答案是——加减味数可以数出来，通常占全方的三分之一到一半；而针对的是什么特征，没有字段；另一位同证病人能不能用，几乎一定是不能，而这个“不能”没有任何地方记录。

问一份精密装配的工艺文件：答案是——不能互换，因为有组号；而组号打在零件上、写在装配卡上、进入备件系统。这里不但有账，账还是强制的：没有组号的零件不许上线。

问一份输血记录：答案是——血型相同仍然不能直接输，必须做交叉配血，而交叉配血有独立的申请单、独立的操作规程、独立的结果记录、独立的责任签字。分类（定型）与配对（交叉配血）是两个步骤、两份记录，而且第二步不能省。这是医学内部把这两件事分开的最成熟的建制，也是本章的直接模板。

问一份器官移植的配型档案：答案是——血型与器官类型只是入门条件，配型分数是另一套独立的系统，它计算的正是“这一个供体与这一个受体之间的位

置关系”，而不是双方各自的类别属性。这里的配对信息不但有记录，还有一个全国性的系统在维护它。

问一份岗位说明书与用人记录：答案是——同一岗位的两个人不可互换是常识，而没有任何字段记录“此人与此岗之间的适配位置”。岗位说明书记的是职级与职责（格），不记这个人在这个格里的位置。这一处与中医的情形几乎完全相同。

五个答案互不相同。其中至少两个不是从主张推出来的，而是查出来的：交叉配血的独立记录、移植配型的独立系统。本章在提出这条问话时并不知道这两处已经把该做的做完了。

八·配格率：一个跨领域量纲不同而比率相同的读数

有了问话，就可以有数。

配格率 = 为使这一次配得上而做的调整量 ÷ 该次处置的总量。

它是一个纯比值，没有量纲。它数的是：**在归类已经完成之后，还有多大比例的处置内容是为了对上这一个个体而添加的。**

在中医：分子是处方中基础方之外的加减药味数，分母是全方药味数。取材是病历与处方，由方剂学判读基础方，判读一致性远高于辨证本身。一张四君子汤加黄芪、升麻、陈皮、半夏的方，配格率是四比八。

在工程学：分子是一个装配中必须按组配对或成对使用的配合副数，分母是配合副总数。取材是装配工艺文件与零件图，逐项可数。

在管理学：分子是一套标准流程在某个单位落地时的例外条款与本地化调整条款数，分母是总条款数。取材是流程文件与其本地版本的差异清单。

在细胞学：分子是同一细胞类型在移植或替换时必须额外匹配的位置特异性或供体特异性标记数，分母是该类型的定义性标记数。成纤维细胞的位置记忆是一个现成的例子——同一类型而不同部位的成纤维细胞携带不同的位置编码，移植后仍保持原部位的身份。

四处量纲完全不同——药味、配合副、条款、标记——而比率的含义完全相同。

八之一·它与既有指标结构性反向

一个新读数最容易犯的错误，是与某个既有主指标高度相关，于是被读成那个指标的代理。检验方法只有一个：**举出一个“既有指标最好看、而这个数最难看”的真实例子。**

本章能举出四个，方向一致。

中医：一份完全遵照诊疗指南、使用标准方、零加减的处方——**指南依从性满分，规范化评价满分，而配格率为零。**这份处方在所有现行指标上都是模范，而按本章的读法，它恰恰是配对信息被完全删除的那一份。反过来，一位老中医的方加减占到一半，在“方证相符度”这类评价上最难看，配格率却最高。

工程学：一条完全互换、零选配的装配线，装配成本最低、备件管理最简单、所有指标漂亮——而它的配合精度上限被单件公差带宽度锁死。**要突破这个上限，必须引入选配，而引入选配之后所有那些漂亮指标都会变难看。**

管理学：一套在全部分部零例外执行的标准流程，合规率百分之百——而它同时意味着没有任何一个分部的特殊情况被承认过。合规率与配格率在这里几乎是数学上的互补。

细胞学：一种被认为高度同质、可大规模扩增、批间差异极小的细胞制品——质量控制指标最漂亮——而它在移植后往往表现出最弱的位置适应性。同质是生产指标，适配是功能指标，两者不同向。

四个例子的方向一致：配格率与既有主指标不但正交，而且在多数情形下**结构性反向**——既有指标越好看，这个数越小，而这个数越小意味着配对信息越少。这使它不可能被读成任何现有指标的代理。

八之二·它自带一件伦理，而且比通常更难写

一个比率一旦进入制度，几乎必然被拿去考核人。因此必须把限制写死。而配格率有一个前两条读数没有的麻烦：**它没有一个明确的好方向。**

第十三节会说明，工程学的证据表明高配格率不是好事——选配是加工精度不足的补偿，工业史的方向是消除它。而本章的分析表明低配格率也不是好事——它意味着配对信息被删除。**两个方向都能被论证成坏，因此这个数一旦进入考核，无论朝哪个方向考核都是错的。**

三条限制因此必须写得更死。

第一，这个读数指的是记录结构，不是人。配格率高，说的是“这一次处置里有多大比例是为对上个体而做的”，不是“这位医生随意加减”或“这位医生

辨证不准”。同一位医生在不同病种上的配格率会差很多。任何把它换算为医师个人评价的做法都是误用。

第二，不得把配格率设为考核指标——无论正向还是负向。这一条是本章写得最硬的一条，而且它与前两条读数不同。理由已在上边说清：这个数没有好方向。它的用途只有一个——**作为疗效评价的分层依据**：配格率显著不同的病例不应被直接合并计算疗效。这是它唯一正当的用法。

第三，已经按这个读数做出不利安排的，必须公开判读规则并给出复核通道。配格率的计算依赖“什么算基础方”这个判读，而这个判读不唯一——同一张方按不同的方剂学传统可以拆出不同的基础方。任何用它做出不利决定的机构，必须公开它用的拆分规则，并允许被评价者提出另一种拆法。

三条缺一，这个读数就不该被使用。

九·回到题目：何谓一个「证」

前面八节做的是地基。这一节回答题目。

九之一·证是一个格，不是一个位

分类给出的是**归属**。它回答“这个属于哪一类”。它不回答“这一个和那一个能不能互相顶替”。

在多数场合，这两件事被绑在一起，因为分类本来就是为可替换而设计的——工程学的公差带、化学试剂的规格、血型的分型，都是如此。绑得住的条件是明确的：**处置手段对类内残余差异不敏感**。一个零件的尺寸只要在带内，装配就不受影响；一袋A型血只要定型正确，绝大多数情形下就可以输。

中医的情形不满足这个条件，而且不满足得相当彻底。一张复方对个体偏性的敏感度，与人群中个体偏性的连续分布，两者的量级不匹配。**在这样一对手段与对象之间，分类给出的归属与使用所需的可替换性之间必然留有缺口。**

于是有了本章的正面回答：

一个「证」是一个格，不是一个位。它把这个人放进一个格里，而这一次治疗成不成，还取决于他在这个格里的哪个位置——而那个位置，证名不承载。

这句话听起来像是在贬低证。它不是。它恰恰解释了证为什么两千年来既不可或缺又永远不够——因为它做的是分类这件事，而分类这件事本来就只做一半。要求它做完全部，是要求它变成另一样东西。

九之二·辨证论治这四个字里有两个动作

“辨证论治”通常被读成一个连贯的过程：辨清证，然后依证立法处方。这个读法把它当作一个动作的两个阶段。

本章的读法是：**这里有两个不同性质的动作，而且第二个不能从第一个推出来。**

第一个动作是分类：把这个人放进一个格。它是可教的、可标准化的、可以写进教材与指南的，因为它处理的是格与格之间的边界。三十年的证候标准化，做的全部是这一件事，而且做得相当成功——**辨证的可传授性从来不是中医的短板。**

第二个动作是配对：在格已经定下之后，决定这一张方与这一个人配不配得上。它处理的是格内的位置。它在记录上的全部痕迹就是那几味加减，而加减的依据不进入任何字段。

“方证相应”这个说法，把两个动作压成了一个。它字面上说的是方与证相应——而临床上真正发生的是方与人相应，证只是中间的一个关口。这一步压缩在语言里不留痕迹，在记录里也不留痕迹，于是配对这个动作从未获得独立的位置。

由此可以说清一件长期说不清的事：**中医传承之难，难在哪里。**

通行的说法是“经验难以言传”，或“悟性”。本章给出一个更具体的定位：**可传的是格，难传的是位。**教材、指南、证候标准传的全部是分类——哪些条目构成哪个证，这一部分传得相当好。而配对这一部分从未进入任何可传的载体：它藏在加减里，而加减在医案中被记录为结果，不被记录为依据。**跟师抄方之所以不可替代，正是因为抄方是唯一能观察到大量配对实例的场合——而观察到的仍然只是结果，依据要靠学生自己反推。**

这也给出一条与通行看法相反的推论：**名医与普通医生的差距，主要不在辨证准，在配格准。**辨证是可教的，教了三十年，差距在缩小；配格没有载体，因而差距不缩小。这条推论可检验：**若把一批名医与普通医生的病历按证型比对，两组的辨证一致性差距应当小于两组的加减内容差距。**

九之三·中医为什么难标准化——一个与流行说法不同的回答

流行的说法有两种：一种说中医不科学所以难标准化；一种说中医是复杂系统、个体化医学，标准化是错误的目标。

本章给出第三种，而且它是结构性的：

现代标准化制度是为分类设计的，它的目标是互换性。而中医临床的核心操作有一半是配对，配对不产生互换性。把一个以配对为核心的实践放进一个为互换性设计的制度里，必然要么被判为不合格，要么被迫删掉配对那一半。

这个回答有几个可以直接检验的后果。

第一，中成药与协定处方的疗效应当系统性低于同证的个体化处方，而且差距随病情复杂度上升。因为中成药把配格率强制为零。这一条在临床上被广泛体感到，而它通常被解释为“中成药力弱”或“药材质量”，本章给出的是另一个机制。

第二，随机对照试验中固定处方的设计，会系统性低估中医的疗效上限，而不是高估。因为它删除的正是配对那一半。这与通常的担心方向相反——通常担心的是个体化带来偏倚，本章指出固定处方带来的是系统性低估。

第三，中医的可标准化程度应当与病种的“处置敏感度”负相关。在处置手段对个体差异不敏感的病种（外治、某些急症的定法），标准化会成功；在敏感度高的病种（内科杂病、慢性病），标准化会一直失败。这不是水平问题，是结构问题。

九之四·与本系列同题另两章的分界

本系列另有两篇处理同一道题的论文，三篇合起来构成三个不同的切面，而且三个切面互相正交。

《让层》（第一章）切的是一次判定内部：同时认定成立的若干层里，动了哪一层、让开哪几层。《吞差》（第二章）切的是同一个人的跨次：一串判定之间跨了几档、登记了几次改变。本章切的是跨个体：两个被判为同证的人之间，方能不能互换。

一次之内、时间之上、个体之间——三个方向，三个不设字段的东西，三个比率。

三篇对同一份材料给出不同评价，这是它们必须并读的理由。取一份“同证、零加减、完全按标准方”的处方记录：《让层》可能判它让层比为零、记录完整；《吞差》在单次上读不出任何问题；本章判它配格率为零，即配对信息被完全删除，风险最高。三个读数互相独立，而一份处方可以同时是三者中的任意组合。

三篇联合起来给出一条没有任何单篇能给出的判断：让层比为零、吞差率高、配格率为零的档案，是最危险的一种——它一次只认一层、跨次不登记变化、且删掉了全部个体化信息，而它在现行的每一项评价指标上都是模范。

九之五·三个词的位置

位置	在本章的结构里是什么	在账上的形态
症	归格所用的输入条目	四诊记录，已有完整字段
证	一个格，给出归属而不给出可替换性	病历“证候”栏，已有字段
配格	格内的位置，决定这一次配得上配不上	无字段（痕迹在加减里，依据无处可记）
方	配对的产物，被当作分类的产物记录	处方栏，有字段而语义错位

这张表里有一行完全空白，还有一行的语义是错位的。处方被当作“这个证的治法”记录，而它实际上是“这个证加上这个人的位置”的产物。这个错位使得任何按证名合并处方的研究，都在合并两样不同的东西。

十·在别的地方，这个缺口长什么样

一个跨领域的判断，只有能在别的领域里据以预测一件具体的事，才算兑现。本节逐处兑现，共十四处。

一·工程学：分组选配的组数是配格率的直接读数。预测：在同一类配合件上，随着单件加工精度提高，选配组数应当下降而配合精度不变——即配格率与加工精度负相关，而与配合要求正相关。这在制造业的工艺演进史上可以直接查，而且它给出中医的一个类比推论：若四诊的分辨能力提高（加入定量手段），配格率应当下降。第十五节把它列为可证伪点。

二·工程学：配对件的“不得单换”是一条强制记账规则。 预测：把“不得单独替换”这类规则移植到中医（某些方的加减不得由他人套用），需要的载体不是禁令而是编号——工程学的做法是给配对件打同一个号。**对应的做法是给“这一张方”与“这一个人”一个共同的标识，而不是给方一个通用名。**

三·工程学：统计公差分析给出了第三条路。 不提高精度、不做选配，而是重新分配各环节的公差，使装配后的累积偏差落在要求内。对应的推论：**中医或许可以通过重新分配“哪些差异由证名承担、哪些由加减承担”来减少配格率，而这需要先把两者分开记账——而这正是本章要求的字段。**

四·输血医学：血型分类与交叉配血是两个独立步骤。 预测：任何只做分类不做配对的替换体系，其失败率不为零，且失败在分类层面看不出原因。这在输血史上有确凿的证据链——同型输血仍有溶血反应，正是这些反应逼出了交叉配血。这条历史是本章最有力的一处兑现，因为它是一次已经完成的、有代价的、可查的实验。

五·器官移植：配型分数是一个纯粹的“位置关系”量。 它不描述供体，也不描述受体，只描述两者之间。预测：任何试图把配型分数拆解为供体属性与受体属性之和的做法都会损失信息——因为它本质上是一个二元关系，不是两个一元属性的函数。对应到中医：**配格不可能被拆成“证的属性”加“人的属性”，因而任何试图靠增加病人分型来消除它的方案都有一个原则上的上限。**

六·细胞学：位置记忆证明同类型细胞不可任意互换。 同一类型的成纤维细胞携带不同部位的位置编码，移植后仍保持原部位身份。预测：细胞治疗中若忽略供体细胞的原始位置，功能整合会系统性变差，而这个差异不能通过提高细胞纯度来解决——**纯度是分类指标，位置是配对指标。**

七·细胞学：调控核心的同源判据本身不保证功能可替换。 两个共享调控核心因而同属一型的细胞，在不同微环境中的功能可以完全不同。这是细胞学自己的一处内部张力，它与本章的主张同构：**同一性判据与可替换性不是一回事。**

八·管理学：例外条款的数目是流程落地时的配格率。 预测：一套零例外执行的标准流程，其在各单位的实际执行偏离度反而更高——因为需求没有消失，只是转入了记录之外。这可以在同一集团的多个分部间直接检验。

九·管理学：关键人依赖是人岗配格的极端形态。 一个岗位由某个特定的人承担而不可替换，通常被当作组织风险来管理，处方是“知识沉淀”与“轮

岗”。预测：这类处方在配格率高的岗位上会失败，因为沉淀的是格（职责与流程），失掉的是位（此人与此岗的适配）。这与本章关于中医传承的分析同构。

十·药学：治疗药物监测与个体化给药方案。同一诊断、同一药物、按体重给药仍然血药浓度差异巨大，因而发展出监测与剂量调整。这是现代医学内部承认配格并为之建账的一处——剂量调整记录是独立的。预测：凡是建立了监测体系的药物，其“同诊断同剂量”的假定被放弃的程度，与该药的治疗窗宽窄成反比。

十一·精准医学：分层是把格切细，不是做配对。这是本章最容易被混同的一处，第十一节会正面处理。预测：分层医学在处置手段对残余差异不敏感的病种上会成功（某些靶向治疗），在敏感度高的病种上会遇到与中医相同的天花板。若分层医学在所有病种上都能通过切细解决问题，本章的核心论证错。

十二·教育：分层教学与个别化教学是两件事。分层是把学生归入若干水平组（格），个别化是针对这一个学生的具体位置调整（位）。预测：只做分层不做个别化的班级，其组内成绩离散度不随分层精细度提高而下降到某个水平之下——这个下限就是配格所在。

十三·服装制造：号型与定制。标准号型是格，量体裁衣是位。这个领域已经把两者的经济学算清楚了：大规模定制的整套方法论就是研究“格切到多细才划算、剩下的用什么方式配”。预测：任何一个领域在处理这个缺口时，最终都会收敛到“标准格+有限调整”的结构，而不是“无限细分”或“全定制”——中医的“基础方+加减”本来就是这个结构，只是它从未被这样命名过。

十四·语言：词典义与用法。一个词的词典义是格，它在这一句里的确切所指是位。预测：任何试图用穷尽义项来消除歧义的词典，其义项数会随语料规模单调增长而不收敛——这与“把证分得更细”面临同一个结构。

十四处里，第四、第十、第十三处给出的是已经完成的建制，可直接作为移植模板；第一、第三、第十一处反过来限制了本章的主张。

十一·与最近的几种说法的分界

一个新提出的东西最大的风险不是被反驳，是被吸收。本节逐一处理最容易发生这种吸收的十种说法。

可替代性（经济学）。把本章的主张剥掉全部中医名词，只留形式，它是这样的：一次通过的归类不保证在使用处可替换，而“能不能替换”所需的信息在归类时被判为零。这个纯形式结构在上游最正式的名字是可替代性——经济学早就区分了可替代物（货币、原油、标准件）与不可替代物（土地、艺术品、二手车）。本章必须正面处理它。分离线在于它是一个关于物类的先验二分，本章说的是同一个类别内部的部分可替换：经济学问“这一类东西是不是可替代的”，答案落在类上；本章问的是“这一类里的这一个与那一个能不能换”，答案落在个体位置上，而且这个位置被记为零。若观察到：把「证」整体判为不可替代物之后，同证同方的疗效差异问题得到处理，则本章被它吸收；若这个判定只是取消了讨论而没有给出任何可清点的东西，则本章提供了它不提供的东西。

可读性与地方性知识（Scott 1998，政治与社会理论）。本章的形状属于“分类越成功越掩盖差异”这一族，而这一族有一个影响极大的现成论述：为了可管理而把复杂现实简化为可读的类别，从而摧毁了那些无法被编码的地方性实践知识。这一位与本章距离极近，必须点名判定。分离线在于它给的是一个关于国家与地方知识的政治经济学论断，本章给的是一个可清点的比率与一个具体的字段。前者的处方是“保留地方自主”，后者的处方是“在分类之后加一个配对步骤并记录它”——而输血医学已经证明，这个步骤可以被完全建制化，它不需要以放弃分类为代价。若观察到：为配对建账之后，配对信息仍然无法被记录（因为它在原则上不可编码），则这一位对而本章错；若配对依据可以被有效记录（如交叉配血结果、剂量调整记录、组号），则本章对。这一条的检验其实已经完成了三次——组号、交叉配血、治疗药物监测都是可编码的配对信息。

精准医学与分层医学。这是本章最危险的一位通俗占位者，因为它的处方看起来与本章要解决的是同一个问题。它说到这一步：通过分子标志物把患者分成更精细的亚组，使每一亚组内部同质，从而实现同组同治。分离线在于它的整个方案是把格切细，而本章论证的是切细有原则上的上限。上限的位置由处置手段对残余差异的敏感度决定，与切分技术无关。若观察到：某个病种通过分层达到了组内可互换（同层的方案可以任意互换而疗效无差异），则在那个病种上分层成功、本章不适用；若观察到分层到某个粒度之后组内差异不再下降而仍需个体调整，则本章对。这个观察在肿瘤靶向治疗领域已经有大量数据可查，两个方向的例子都存在——这恰恰是本章所预测的（敏感度不同的病种结果不同）。

因人制宜与三因制宜（《素问·异法方宜论》等，中医内部）。它说到这一步：治疗要考虑时、地、人，同病异治，因人而异。这是中医内部与本章最近的说法，而且早了两千年。分离线在于它是一条医嘱层面的原则，本章是一个记账层面的空缺：因人制宜说了要考虑人，但从未要求记下“这一次考虑了这个人的哪一点”。若观察到：遵循因人制宜的医案中，能够系统地反推出加减的依据，则依据已被记录，本章的空缺主张为伪；若医案只能给出结果（这张方）而反推不出依据，则本章对。这一条可以在现存医案上直接检验，而且医案的体例本身就是答案——医案记的是脉案、证名、处方、效果，独独没有“为何加这一味”的字段。

方证相应（黄煌等）。它说到这一步：方与证之间存在直接的对应关系，识证即可用方，不必经由病机推演。这一位与本章的关系微妙——它实际上是在往本章的方向走（它把“人”的特征如体型、面色、腹证纳入了识别），但它把这些特征仍然编码为「证」的一部分。分离线在于它把配对信息塞回了分类，本章主张两者分账。若观察到：方证相应体系发展到成熟时，其“方证”的数目趋于收敛（是一个有限的谱），则它成功地把配对转化为了分类，本章这一处错；若观察到方证数目随实践积累持续增长而不收敛，则它遇到的正是本章所说的上限。这一条可以在这一学派的历代著作上直接数。

隐性知识（Polanyi 1966）。它说到这一步：我们知道的比我们能说出来的多；某些技能只能通过实践与师承传递。这是“传格而失位”这个形状最著名的名字，必须点名。分离线在于它把不可言传归因于知识的性质，本章归因于记录结构的缺失——两者给出相反的处方。隐性知识的处方是延长师承、增加实践；本章的处方是加一个字段。若观察到：为配对依据建立字段之后，传承效率没有改善（因为依据确实说不出来），则这一位对；若观察到建账之后可传部分显著增加，则本章对。三个已有的建制再次构成证据：组号、交叉配血结果、血药浓度调整记录，全都是曾被认为“要靠经验”而后被成功编码的东西。

自然类与名义类（哲学）。它说到这一步：有些类别是世界本身的关节，有些是我们为方便而划的。分离线在于这个二分处理的是类别的地位，本章处理的是类别与使用之间的关系——一个类别可以是完全自然的（血型就是），而它仍然不保证可互换。血型是本章最好的一个反例：它是一个有确凿分子基础的自然类，而同型仍需交叉配血。这条例子同时否掉了“只要找到自然类就能保证互换”这个直觉。

分组选配与配对件（工程学）。这是本章的推翻材料，同时也是最近的实现形态。分离线在于**工程学为它建了账而且账是强制的（无组号不得上线）**，中医没有。本章不能声称这个问题无人处理过——本章只能声称它在中医的证方关系上未被处理。这一位同时是第十七节赌注的形态来源之一。

交叉配血与移植配型（输血与移植医学）。这是本章所要求的字段在医学内部的完整实现：分类与配对是两个步骤、两份记录、两次签字，且第二步不可省。本章因此不能声称这是一个新发现。本章能声称的是：**这套在输血上被血的教训逼出来的建制，在中医的证方关系上从未被移植过，而两者的结构完全相同。**这一位是本章的最强模板，第十七节的赌注照它写。

兼夹证与证素辨证（中医内部）。它说到这一步：通过增加证的维度与组合，逼近个体的实际状态。分离线在于**它沿着切细这条路走，本章论证这条路有上限。若观察到：证素维度增加到某个数目后，同证素组合的病例之间处方可互换，则切细成功、本章错；若观察到证素组合数已接近病例数而处方仍不可互换，则本章对。**这一条可以在现有的证素辨证数据库上直接算——**组合数与病例数的比值，本身就是一个可以立即查的读数。**

十条之中，最近的一位是交叉配血——它已经把本章要的东西完整地做出来了。本章之所以仍不被它吸收，是因为交叉配血处理的是一次二元反应的相容性（会不会溶血），是一个二值判定；而配格是一个连续的位置，它决定的不是“能不能用”而是“用多少、加什么”。二值的相容性判定可以建成一个通过／不通过的关卡，连续的位置判定不能——这正是中医这一侧至今没有对应建制的技术原因，而不只是意识问题。这一条本章必须写出来，否则是掩饰。

十一之一·这些说法共同看不见的那一块

逐个问最近的几位“它把什么当作给定，因此结构性地看不见什么”，答案有一个共同的形状。

精准医学把“组内差异可以通过继续切分消除”当作给定，因此看不见处置敏感度设定的上限；一旦承认这个上限，它的整个路线图就需要一个终点，而它目前没有终点的概念。证素辨证把“证的维度可以逼近个体”当作给定，因此看不见组合数逼近病例数时分类就消失了；承认它，证素的意义就要从“逼近个体”改写为“提供一个有限的格”。隐性知识把“不可言传源于知识的性质”当作给定，因此看不见有些不可言传只是没有字段；承认它，这个领域的处方要从

延长师承改为改造记录。因人制宜把“考虑了就够了”当作给定，因此看不见考虑的内容需要一个载体才能传下去。

四条盲区背后是同一块地：

一个分类系统的任务，在它给出归属的那一刻就完成了。

归属即完成，这是所有人共同站着的那一块。分类学在它上面，标准化在它上面，指南在它上面，绝大部分评价体系在它上面。“**未被任何学科辨识**”的真正来源不是没人看，是看的人都站在它上面。而那三个把配对建制化了的领域——精密装配、输血、治疗药物监测——恰恰都是被事故逼出来的：装不上、溶血、中毒。**它们不是想明白的，是撞出来的。**

本章这一次只掀开了它的一角——只处理了“分类之后还有一步”，没有处理“这一步能不能被切分成可传的单元”。第十六节交代这一角。

十二·与本系列各章的分界

本系列此前已有三篇使用了完全相同的三个学科，另有两篇处理完全相同的题目。必须逐篇指名。

《叠裕》（《治未病立账学》第三章）（何谓未病，同样取材于细胞、工程、管理）。那一篇处理的是各单位全部合格而这些合格依据同一份不可分割的储备重复整份计入；读数是叠计比。分离线在于**方向**：那一篇是多个单位共用一份余量（分母被重复使用），本章是一个类别内部的位置未被记录（分子与实际所需不匹配）。三家的立场在两篇里全部不同：那一篇的细胞学占邻域相对判定、工程学占损伤容限、管理学占已出现的结果；本章的细胞学占谱系与调控核心、工程学占互换性与公差、管理学占共享计量制度。同一份“**四科规范用药**”的档案，那一篇查各科声明的余量加起来是几倍，本章查每一科的方能不能给同证的另一个人用。

《漏汰》（《治未病立账学》第六章）（未病如何被识别，同三学科）。那一篇处理的是筛选环节中被漏掉的那一部分与参照的脱钩。分离线在于**环节**：那一篇在识别的入口（该进来的没进来），本章在归类的出口（进来了、归对了、而用不上）。若一个问题在筛选环节完全正确执行之后仍然出现，它不属于那一篇。

《功债》（《治未病立账学》第九章）（治未病为何实践最少，同三学科）。那一篇是关于动力的，处理的是成功的预防所留下的、必须一直被供养的残留；读数是续养率。分离线在于题型：那一篇答的是为什么，本章答的是是什么。两者可以在同一份档案上同时成立且互不相干——一个续养率高的病人，其配格率可以是任意值。

《让层》与《吞差》（问题：何谓一个证）。第九节之四已经给出三个切面的分工：一次之内、时间之上、个体之间。此处补一条判决性对照：一份“同证、零加减、二十次不变”的档案，《让层》判让层比为零、《吞差》判吞差率可能为一、本章判配格率为零。三个读数在同一份材料上给出三个不同方向的结论，而三者都为真时，这份档案在现行的每一项评价指标上都是模范。这是三篇合起来才能说出的一句话。

《撑平》（《治未病立账学》第一章）《阙前距》（《治未病立账学》第二章）《悬量》（《治未病立账学》第八章）《欠返》（《治未病立账学》第五章）（治未病与未病诸题）。与本章的共同分界很简单：它们全部处理同一个对象在时间上的读数结构，本章处理不同对象之间的可替换关系。若一个现象在只观察一个病人时仍然出现，它不属于本章——本章要求至少两个个体，因为可互换性是一个二元关系。这条判据可以在任何具体案例上直接执行。

本系列这些篇目共用一个未言明的做法：都在找各领域账本上被当作垃圾、余数或噪声的那一格。这是共同的强处，也是共同的弱点——第十六节交代它在本章身上的形状。

十三·三处反过来削弱本章的地方

若三个来源全都顺着一条判断说，那它多半只是在这三处各挑了一句合用的话。本章有三处削弱，其中第一处动的是价值排序，第二处动的是适用范围。

第一处，工程学削掉了本章的价值排序。本章原本会写下这样一句更强的话：**配对是比分类更根本的操作，个体化优于标准化。**工程学不允许。

分组选配在工程学里不是一个值得追求的东西，它是一个**不得已的补偿**。它带来真实的成本：分组、打号、分箱存放、备件必须按组供应、维修时必须找到同组件。整个工业史的方向是**消除选配**——提高加工精度、改进设计使配合不那么敏感、用统计公差分析重新分配各环节的余量。互换性带来的收益（备件通用、大规模生产、任何人都能维修）巨大到无法放弃。

若配格率高是好事，那么最好的机器应当是每一处都要配研的机器。事实恰恰相反。

本章因此不主张个体化优于标准化，也不主张配格率越高越好。本章剩下的主张只有一句：**这个数应当被记下来，而它现在没有被记下来**。第八节之二那条“不得把配格率设为考核指标——无论正向还是负向”，正是从这里来的。

工程学还给出了本章原本没有看见的第三条路：**统计公差分析**。不提高精度、不做选配，而是重新分配各环节的公差，使累积偏差落在要求内。移植过来就是：**重新分配“哪些差异由证名承担、哪些由加减承担”**。这条路本章原来不知道，而它是三条路里最省的一条——但它有一个前提：**两者必须先分开记账**。这一处削弱因而同时加强了本章的核心要求。

第二处，精准医学与细胞学一起削掉了本章的适用范围。本章原本会写下：**任何分类系统都不可能保证类内可互换**。这句话太强了。

反例是明确的：血型分类在绝大多数情形下确实保证了可互换（交叉配血的阳性率很低）；某些靶向治疗的分子分型确实做到了同型同治；工程学的绝大多数零件确实完全互换。**这些都是分类成功兑现了可替换性的真实案例**。

差别在哪里？在**处置手段对类内残余差异的敏感度**。零件的装配对带内差异不敏感，所以互换成立；复方对个体偏性极其敏感，所以互换不成立。

本章因此把适用范围收窄为：**处置手段对类内残余差异敏感、且个体差异连续多维的那一类系统**。这个范围比“任何分类系统”窄得多，而且它是一个可以事先判断的条件——问一句“把这个处置原样给同类的另一个对象，结果会不会显著不同”就知道。

细胞学在这里补了一刀：它主张同一性由发生史裁定，而这意味着**存在一个自然的裁决者**。若中医的证也有一个自然的裁决者——某个共同的病理发生机制——那么“同证不可换”就只是分类还没做到位，本章的主张会退化为“标准还不够细”。本章必须承认这个可能性：**本章的立论依赖于一个经验假定——在四诊这套采集手段下，不存在能保证互换的自然分类**。这个假定可以被推翻。若某天生物标志物给出了保证互换的证候分类，本章失效。这不是一个遥远的假想，它正是当前证候客观化研究的目标。

第三处，管理学削掉了本章关于“建账即可”的乐观。本章原本会写：为配格建立字段，配对信息就可以被传递。管理学的经验不支持这种乐观。

管理学早就知道：一个字段一旦被建立并进入考核，它记录的内容会立刻退化为关于该字段的应付。例外条款一旦要求申报并统计，申报的内容会迅速变成一套模板化的理由；关键人依赖一旦被要求登记，登记的是岗位而不是那个人真正不可替代的地方。配格依据一旦被要求填写，很可能变成一栏“因人制宜”四个字。

本章接受这一处，并且把它写进第八节之二的第二条限制（不得设为考核指标）与第十六节的反噬预言。这一处削弱使本章的处方从“建一个字段”降级为“建一个字段并使它不进入考核”，而后者在制度上远比前者难。

三处削弱之后剩下的主张，比本章最初打算写的窄得多，也弱得多。这是应该的。

十四·不适用的边界

本章的判断在下列情形不适用，或需要重大修改：

处置手段对类内残余差异不敏感的场所。 见上一节第二处。此时分类确实兑现了可替换性，配格率趋近于零而这并非缺失。

只有一个个体的场所。 可互换性是一个二元关系，至少需要两个对象。本章对单个病人的纵向问题没有内容，那属于本系列另外两篇。

处置不可分割的场所。 本章的读数依赖于把处置拆成“基础”与“调整”两部分。若一次处置在原则上不可拆分（一次手术、一次不可调整的操作），则分子无法界定。针灸的取穴与手法在多大程度上可拆，本章没有把握。

基础方判读不唯一而争议极大的场所。 某些方剂的源流复杂，什么算基础方本身就有多种传统。此时配格率的计算需要先声明拆分规则，而不同规则给出的数不可直接比较。

配格依据在原则上不可编码的场所。 若某些配对判断确实无法被表述（第十一节隐性知识那一位所主张的），本章的处方无效。本章认为三个已有建制构成反证，但这并非决定性的——那三处编码的都是二值或标量，配格是一个高维位置。

急性、单次、目标明确的处置。 此时格与位的差别不重要，因为处置的时间窗不允许调整。

十五·可以让本章翻车的观测

本章的主张可以被分层引用。**第一层是配格这个构念**，最强也最易倒。**第二层是那张四格表**，即使构念被推翻，“同证／可换”这个二维分类仍可能有用。**第三层是一条不依赖前两层的经验主张**：按证型合并计算疗效的研究，其合并前提（同证内可互换）从未被检验过。这一层最容易检验，且检验它不需要接受本章的任何概念。

十五之一·最强的竞争解释，以及排除它的条件

必须先写出那个最朴素、最难反驳的说法，不能挑一个好打的。它是：

其实不就是辨证还不够细吗？

这个解释极强：它符合所有人的直觉，它有三十年的研究方向作为背书，而且它的处方明确可执行——继续细分。

排除它的条件写成这个形状：

控制辨证细度（以证素维度数或亚型层级数为自变量）之后，若“同证病例之间处方不可互换的比例”随细度提高而单调下降并趋近于零，则本判断为伪；若该比例随细度提高而下降至某个明显大于零的水平后不再下降，且该水平的高低与病种的处置敏感度呈剂量—反应关系，则本机制成立。

关键在于两个说法预测的曲线形状不同：

- 若只是不够细，曲线应当**趋近于零**。
- 本章预测曲线**收敛到一个大于零的平台**，且平台高度由处置敏感度决定而非由分类技术决定。

这是一条形状预测，比水平预测难被巧合满足。而且它可以在现有数据上做——证素辨证的数据库里，证素组合数与病例数的比值是现成的。

十五之二·样本纪律

凡涉及剂量—反应关系的检验，样本至少六个单元，且在关键混淆变量上同质。**本章明确排除跨国或跨地区的两两比较作为证伪检验。**

可用的同质设计：同一家中医院的六个以上专科（自变量是病种的处置敏感度，这在同一制度环境内有真实变异）；同一专科内按证候标准细度分层的多个时期（标准修订前后）；同一批医师在不同病种上的加减率比较。

十五之三·其余可以让本章翻车的观测

一·若加入体质分型之后，同证同体质的两位病人处方可以互换，则配格被体质吸收，本章错。

二·若辨证水平越高的医生加减越少，则加减是辨证不准的症状而非配对做得细的症状，第五节之一那一条为伪。

三·若同证病例处方的加减部分重合度不低于基础方部分重合度，则第五节之三的推论为伪。

四·若四诊加入定量手段之后配格率显著下降并趋近于零，则第十节第一处成立而本章的核心论证被架空——**这一条是本章最愿意被证伪的一条，因为它被证伪意味着证候客观化成功了。**

五·若方证相应体系的方证数目随实践积累而收敛到一个有限的谱，则它成功地把配对转化为了分类，本章这一处错。

六·若为配格依据建立字段之后，其填写内容在两年内没有退化为模板化套话，则第十三节第三处的管理学削弱不成立，本章的处方可以更乐观。

七（正向读数，写成顺序而非相关）·若引入配格记录，那么“**同证病例被拆分计算疗效**”的做法出现，应当早于“**证候标准被修订以纳入配对维度**”。记账在先，标准修订在后。若观察到标准先动而记账未动，则说明配对又一次被塞回了分类——这正是本章预测的最可能的失败方式，届时本章关于“两者必须分账”的主张为伪。

八（低成本可实做，所需材料已经存在）·取任意一家中医院一个专科一年的门诊处方，按证型分组，把每张方拆成基础方与加减，算两部分各自的组内重合度。若两个重合度接近，则同证确实同方，配格现象不存在。这个检验不需要采集任何新数据，也不需要任何盲评组织，一个人用两周就能做完。**它是本系列至今给出的成本最低的一条检验。**

若本章被证伪，学到的东西是明确的：那意味着证的分类在原则上可以细到保证互换，因而证候标准化的方向完全正确，只是尚未完成。这会把资源引向继续细分与客观化，而不是引向为配对建账。

十五之四·本章解释不了的两个洞

第一个洞：配格能不能被切分成可传的单元？工程学的组号是一个标量，交叉配血是一个二值，而配格是一个高维位置。**本章不知道一个高维的配对判断不能被有效编码**，也不知道若能，需要多少维。

第二个洞：配格与配格之间有没有结构？两位医生对同一个病人的加减不同，这两个不同是随机的，还是指向同一个位置的两种表达？若是后者，配格是可以被客观化的；若是前者，本章的整套读数会退化成一个关于医生风格的统计。本章给不出答案，而这个答案决定本章的分量。

十六·本章自己的配格率

按本章自己的判据给本章定位，结果不好看，而且必须说出后果。

本系列这一系列篇目，每一篇都用了同一个基础方：找一个被三家账本判为垃圾的对象，造一个二维辨别格，给一个可通约的比率，写一句零情态词的话，加一条写死日期的赌注。**这是基础方**。每一篇不同的地方，是三个学科、一个题目、一个新造词——**这是加减**。

于是可以直接算：

本章的配格率，按药味比的算法，大约是五分之一到四分之一。

后果有三个。

后果一：本章正落在自己指认的那一格里。第六节说那一格的三条签名是短期指标改善、失效不产生信号、越正规化越严重。本系列完全符合：每一篇的基础方使论证结构稳固（短期好看）；若某一篇的加减其实并不适配它那个题目，不会有任何信号，因为读者读到的是一篇结构完整的论文（无信号）；而本系列的方法越成熟、基础方越固定，加减的空间就越小，**同一个基础方被套用到不适配的题目上的风险就越大（越正规化越严重）**。

后果二：本章对本系列的处方，本系列自己没有执行。本章主张为配对依据建立字段。移到本系列就是：每一篇都应当写明“这三个学科为什么适配这个题目，而不是另外三个”。本系列至今没有一篇这样写过——包括本章。三个学科是被指定的，而“为什么这三个适配这道题”这个判断，从未被记录，只在成稿的顺滑程度里被间接体现。**这正是本章所说的：依据藏在结果里。**

后果三：本章无法自证它的加减是适配的。与前两篇不同，这一篇的自我诊断有一个特别难堪的地方——本章第十五节之四第二个洞（配格与配格之间有没有结构）直接适用于本系列自己：本系列各篇的加减不同，究竟是指向不同题目的适配，还是只是三个学科被指定之后的自然变形？本章给不出答案。能给出答案的只有一个动作：把本系列某一篇的三个学科换掉，看结论会不会变。这个动作本章没有做，而它才是真正的检验。

十六之一·反噬预言，且没有出口

配格率一旦进入制度，会发生什么？

若被当作正向指标（加减多说明个体化做得好），最省事的达标法是多加几味无关紧要的药。这不但零成本，而且有真实危害——复方加大，配伍关系变复杂，不良反应风险上升。

若被当作负向指标（加减多说明辨证不准），最省事的达标法是照抄标准方。这同样零成本，而且它直接删除配对信息，正是本章最要防止的事。

两个方向都会被攻破，而且是从同一个地方——分子由医生自己写。加减几味完全在开方人的笔下，没有任何外部约束。

有没有出口？本章找过三条，三条都不通。

第一条是不设为考核指标、只用于疗效分层——这是本章推荐的唯一用法。但管理学的经验（第十三节第三处）说明：一个数一旦存在于系统里并被计算，它进入考核只是时间问题。

第二条是由外部判读加减的必要性——但判读一味药加得必不必要，需要的知识与开方本身相同，成本高于测量本身。

第三条是只记依据不记比率——但不记比率就无法用于疗效分层，而分层是本章唯一主张的用途。

根本的困难在于：这个数没有好方向，而没有好方向的数一旦进入考核，无论朝哪个方向考核都是错的，且都能被零成本地满足。我看不到出口。

十七·一条写死日期的赌注

本章压这一条：

在二〇三七年十二月三十一日之前，中医临床研究的疗效评价规范（或中药新药临床研究的指导原则）中，将出现一条要求：按证型合并计算疗效之前，必须先报告同证病例之间处方的差异程度，且差异程度超过规定阈值的病例组不得直接合并。

这条赌注的形态照着输血医学写的：分类完成之后必须再做一步配对检验，第二步不可省，且第二步有独立记录。输血用了几十年和许多条人命才走到这一步。

以下五条写死什么不算命中：

一·只在方法学讨论或局限性说明中提及“处方存在差异”的，不算命中。这现在就已经在写了。

二·只要求固定处方、禁止加减以保证干预一致的，不算命中，而且构成对本章的否定。这是当前最常见的做法，它通过删除配对来达到可合并，正是本章指认的问题。

三·只报告处方差异而不影响合并规则的，不算命中。报告而不影响决策的字段等于不存在，这一点第五节之一已经论证过。

四·差异程度的计算规则由研究者自定而无统一拆分标准的，不算命中。没有共同的拆分规则，这个数在两项研究之间不可比，而不可比的数无法进入规范。

五·把它设为考核指标（无论正向负向）的，不算命中，而且构成对本章主张的违反。第八节之二第二条与第十六节已经论证：这个数没有好方向。若最终落地的制度把它变成了一个越低越好或越高越好的指标，那么本章不但不算命中，本章的主张被误用了。

移植失败也是一个结论——它会说明中医的证方关系与输血的型血关系在结构上有一处本章没有看见的根本不同。而最可能的那一处，第十一节末尾已经指出：输血的配对判定是二值的，可以建成一个通过／不通过的关卡；配格是连续的高维位置，建不成关卡。若赌注落空，最可能就是落在这里。

十八·结论

本章从一个每天发生而无人当作问题的现象出发：同一个证，同一位医生，两位病人，两张不同的方。

通行的解释是辨证不够细，处方是继续细分。本章论证这条路有一个原则上的上限：**把类分到保证同类可同方，其终点是每个人一个类，而那就不再是分类。**上限的位置不由分类技术决定，由处置手段对类内残余差异的敏感度决定，而中药复方对个体偏性的敏感度极高。

三个互不通约的传统——细胞学的发生史判据、工程学的规格带判据、管理学的共享制度判据——在“该由什么裁定”上势不两立，而在一件事上完全一致：**归为同一类，就意味着在需要它的地方可以互相替换。**分类的意义在于可替换性，这是三家共同的、从未说出口的假定。

推翻它的材料就在其中一家自己手里。工程学的分组选配与配对件证明：两个都落在同一公差带内、检验全部合格的零件，可以不能互换；而工程学为这个差额造了一个强制字段——组号。“合格”与“可互换”在工程学里早已分家，并且分开记账。

由此得到的主张是：一个「证」是一个格，不是一个位。它给出归属而不给出可替换性；决定这一张方与这一个人配不配得上的，是格内的位置。这个位置本章称为配格，它在病历、证候标准、疗效评价三套账上都不设字段——它的全部痕迹是处方里的那几味加减，而加减被当作结果记录，不被当作依据记录。

两根轴给出四个格：同证可换的是真同类，异证不可换的是正常，异证可换的是分类过细——三格都有位置。**只有同证而不可换的那一格，既有后果又无位置**，而且它通过全部形式审查：短期指标上表现为改善（个体化提高疗效，而功劳记在证名上），失效不产生任何信号（无效总被读成辨证不准，从不被读成分类不保证互换），且越规范化越严重（标准化的目标就是消除加减）。

一句不含情态词的问话把它落地：**这一张方里有几味在基础方之外？针对的是这个人的哪一个特征，记在哪个字段？同一时期辨出同一个证的另一位病人，能不能用这一张方？**由此得到一个跨领域量纲不同而比率相同的读数——配格率。它与指南依从性结构性反向：一份零加减的标准方处方，依从性满分而配格率为零。

回到题目：“**辨证论治**”这四个字里有两个不同性质的动作——分类与配对，而第二个不能从第一个推出来；“**方证相应**”这个说法把两者压成了一个；中医传承之难，难在可传的是格、难传的是位；而中医难标准化，不是因为不科学，也不是因为标准化是错误的目标，而是因为现代标准化制度是为分类设计

的、它的产物是互换性，而中医临床的核心操作有一半是配对，配对不产生互换性。

三处削弱把主张收窄了：工程学的工业史证明消除选配才是方向，因而本章不主张个体化优于标准化，只主张这个数应当被记下来；分类成功的真实案例（血型、某些分子分型、绝大多数零件）把适用范围收窄到“处置手段对类内残余差异敏感”的那一类系统，而细胞学进一步指出，本章依赖一个可被推翻的经验假定——在四诊这套手段下不存在保证互换的自然分类；管理学则削掉了“建账即可传递”的乐观，因为一个字段一旦进入考核，它记录的内容会退化为对该字段的应付。

本章与本系列另外两篇同题论文构成三个正交的切面：**一次之内（让层）、时间之上（吞差）、个体之间（配格）**。三个读数互相独立，而三者同时为真的那份档案——一次只认一层、跨次不登记变化、零加減——在现行的每一项评价指标上都是模范。这是三篇合起来才能说出的一句话。

本章自己的配格率约为五分之一到四分之一，本章正落在自己指认的那一格里；而本章对本系列的处方（写明三个学科为何适配这道题），本系列自己从未执行过。至于配格率一旦进入制度会怎样——它没有好方向，因此无论朝哪个方向考核都是错的，而分子完全在开方人笔下。**对此我看不到出口。**

最后是那条赌注，压在二〇三七年十二月三十一日：按证型合并计算疗效之前，必须先报告同证病例之间处方的差异程度，超过阈值不得直接合并。五条不算命中已经写死，其中最要紧的两条是：**固定处方以保证干预一致的做法不算命中且构成对本章的否定；把这个数设为考核指标同样不算命中，而且构成对本章主张的误用。**

第二编 复核

它凭什么被复核

第四章 并身

——一个证的复核为什么不是重测——论已执行的判断如何并入被复核的对象，以及可拆分点落在哪一步

提 要 中医临床有一个从未被正面处理的程序问题：一个证被确立之后，据以处方，服药若干日再诊，此时得到的证与前一版不同——现行的解释只有两种，要么“当初辨错了”，要么“病情变了”。本章指出还有第三种，而它在任何一份病历、任何一套诊断标准、任何一份疗效评价方案里都没有科目：上一版证在它成立的那段时间里，通过它所指导的治法，已经把复核所要面对的那个身体改写了一部分；这一份改变随后被读作身体自身的属性。本章把它命名为**并身**——像子公司并入母公司报表之后不再单独列示。本章的承重命题是：一个证的复核，不是对同一对象的重复测定，不是对同一构念的平行施测，也不是对一条介入档案的再审，而是对一个已被上一版判断改写、且取不到未受影响副本的对象所做的判定；因此复核的效力不取决于它有多独立，而取决于上一版执行的产物有没有被单独立账。材料取自分析化学的方法验证与能力验证、教育测量的效度论证与练习效应、艺术学的归属鉴定与保存修复档案三处成熟范式的互相顶撞，推翻它们共有前提的那件证据来自分析化学自己——在没有标准物质可及时，“算数”的那个终点由当轮参加者的稳健共识重新生成。本章给出一个可在三个领域并排清点的读数——**留底率**：一次复核所依据的证据项中，取自首次处方之前已固定下来、或结构上不受其影响的那部分所占的比例；并给出一句不含情态词的问话：**这一次复核所依据的记录里，有哪几条是在上一次开方之前就已经写下的？** 结论是：复核不能靠加强盲法变得可靠，只能靠在执行之前先把底留下来。

一·一个复核为什么会变成一次新的确立

先摆一个每天都在发生、而没有任何一方觉得需要解释的场景。

一位病人初诊，医生四诊合参，判为某证，开方七剂。七日后复诊，舌象变了，脉象变了，主诉变了三条里的两条。医生重新辨证，得到另一个证名，改方。这一次改方，在病历上留下的痕迹是两行：前一版的证名，与后一版的证名。中间那个箭头，不写理由，因为没有栏位要求写。

如果有人追问那个箭头，回答总是落在两条里的一条：**要么当初辨得不准，要么病情本身移动了**。这两条穷尽了现行的解释空间。前一条把差异记在医生名下，是误；后一条把差异记在疾病名下，是变。两条都能记平账：误可以进入“辨证准确率”，变可以进入“病程转归”。

但把这两条摆在一起看，会发现一件怪事：**它们都假定复核所面对的那个身体，与初诊时那个身体，是同一个可以被反复读取的对象。**误的说法假定身体没变而我看错了；变的说法假定身体自己变了而我这次看对了。两条都不处理第三种可能——

这七天里，病人吃了我开的药。

七剂药下去，汗出了，便通了，睡沉了，舌苔从厚腻转薄，脉从弦转缓。这些变化里有多少是“病自己走到了这一步”，有多少是“我把它推到了这一步”，在四诊信息里没有任何标记可以分开。而下一次辨证，读的正是这些四诊信息。**换句话说，复核所依据的证据，有相当一部分是上一版判断的执行产物。**

这件事在临床上人人知道，在方法上无人处理。人人知道的证据是：任何一个有经验的医生复诊时都会先问“上次的药吃了几剂”“吃完什么反应”，这句问话的功能正是**把执行产物从对象身上剥出来**。无人处理的证据是：这句问话的答案不进入任何一份结构化的病历字段，不进入证候诊断标准的任何一条，不进入证候疗效评价量表的任何一个条目，也不进入任何一项“辨证一致性”研究的分析模型——那些研究测的是不同医生对同一份四诊资料判出同一个证的比例，而那份四诊资料本身是在治疗之中采集的，没有人问过它是第几剂药之后的。

于是问题可以被写成一句话：

一个证如何被确立并被复核，才算数？

这个问句里藏着一个被所有人默认的顺序：先确立，后复核，复核是把确立那一趟重走一遍，看是否到达同一处。本章要说的是，这个顺序在中医的临床对象上不成立，而且不是因为中医不够严格，是因为**被判断的对象与执行判断的场所是同一个**——身体既是判据的来源，又是治法的作用面。凡是这种情形，复核就不可能是重走。

要看清这一点，单在中医内部转是转不出来的：中医没有一套关于“什么叫复核”的现成语法，它只有“观其脉证，知犯何逆，随证治之”这十二个字，而这十二个字恰恰是本章要指认的那一处**被现代验证范式删掉的科目**（第六节）。所以本章从外面借三副眼睛：三门学科各自有一套跑了几十年到几百年的成熟复核范式，三套范式互不兼容，而且冲突不在“谁更严格”，在它们各自把复核的**终点锁在了不同的地方**。

二·三种成熟的复核范式，与它们各自锁死的那个终点

三门学科都要回答同一个问题：一个判断凭什么算数、又凭什么被推翻。三门学科给出的答案不能并存。

二之一·第一家：分析化学——算数落在可复现的读数上

一个测定结果算数，是因为在受控条件下重取一份物料重测，读数落在规定的限内。

这一家的路径是从受控条件出发，经操作序列，落到读数：先把温度、基质、仪器校准、试剂批号这些条件固定下来（条件），再把取样、前处理、进样、积分这些步骤写成不许改动的操作规程（路径），最后看那个数复不复现（落点）。国际标准把这两个层次分得很清：同一实验室同一人同一天做，叫重复性；换实验室换人换设备做，叫复现性；两者之差本身是一个要报告的量。方法验证的那张清单——专属性、线性、准确度、精密性、检出限、耐用性——每一项验的都是“换一个条件，那个数还在不在原处”。

这一家能成立，靠一个不写在清单上的前提：样品可分割、可留样。复核所测的从来不是同一份样品——那一份在第一次测定时已经被消解、被衍生化、被烧掉了。复核测的是同一批物料分装出来的另一份等分样。留样制度、平行样、加标回收，全部建立在“物料可以被切成互不影响的若干份”这一条上。抽掉它，这一家的整套范式立刻失效。

化学人对这一条有多重视，看它的取样理论就知道：一份代表性样品不是被保存下来的，是被制造出来的——破碎、混匀、缩分、再缩分，每一步都有各自的误差项。代表性是被构造的，不是被继承的。这句话本章后面要用。

二之二·第二家：教育测量——算数落在那条推论链上

一个判断算数，不是因为分数复现，是因为从观察到结论到使用的那条推论链，每一环都有证据支持。

这一家在二十世纪后期完成了一次彻底的转向：效度不再是测验的属性，是对分数的解释与使用的属性。据此，一次评价的复核不是重考一遍看分数是否一样，而是把那条推论链拆开逐环检验——评分这一环（打分规则是否被一致执行）、概化这一环（这一次表现能否代表这一类任务上的一般表现）、外推这一

环（考场里的表现能否外推到考场之外）、蕴含这一环（据此做出的安排是否得到支持）。复核就是找出最弱的那一环。

这一家能成立，也靠一个前提：任务可重出。平行卷、等值题库、题目参数库，存在的全部理由是让同一个构念可以被再次施测，而不必重复同一道题。抽掉平行卷，这一家就只剩下重考同一张卷子，而那测的是记忆不是能力。

这一家自己非常清楚“施测会改变被测者”——练习效应、重测效应、测试本身增强记忆，都是它自己几十年反复发表的东西。**它的处置是把这一份当作干扰项，用平行卷控制掉。**这个处置很有效，也正因为有效，它使这一家从来不必把这一份当作“对象已被上一次判断改写”的证据去读。这一条本章第四节要回来用。

二之三·第三家：艺术学——算数落在能否被重新安放回一片共同体的实践条件里

一幅画归在谁名下算数，不是因为某位专家看了一眼，也不是因为某台仪器出了一个数，是因为**三支线索合流**：来源链（这件东西从画室到今天的每一次易手是否可追）、目鉴（共同体内部对手法的判别）、技术检验（X光、红外反射、颜料与基底的断代）。任何一支单独都不足以确立，任何一支单独也不足以推翻。复核在这一家是**重新把这件东西安放回一片实践条件之中**——工坊制度、材料供应、同代作品的比对、赞助与订件的记录。

这一家有一个另两家做不到、也不想做的动作：它的技术检验读的正是历次介入层。X光下能看见底稿被改过几次，红外能看见修补的边界，颜料断代能把十九世纪的清洗与二十世纪的重涂分层剥出来。保存修复档案是一份逐次登记的流水账：哪一年、谁、用什么溶剂、动了哪一块。

而这一家能成立的前提，与另两家正好相反：原物不可分割、不可重出。没有等分样，没有平行卷，只有这一件。既然取不到未被介入过的副本，这一家就干脆**放弃独立性**：鉴定不是盲的，来源链必须被知晓；修复不是从零开始，历次介入必须被读出来。

二之四·三家在打架，而且打得没法调解

把三家摆在一起，冲突立刻显形，而且不是“谁更严格”这种可以判胜负的冲突：

化学与教育学：化学把“上一次执行留下的东西”排除在被测对象之外——换一份就是了；教育学把它算作误差方差要抵消掉——用平行卷就是了。两家给出的归属不相容：一个说它不在样品里，一个说它在被测者身上但可以被平掉。若化学对（另取一份即可），教育学所谓“同一被测者身上的推论链”就没有对象；若必须在同一被测者身上复核，化学的等分样范式在此根本用不上。

化学与艺术学：这一对是最硬的。化学每天在做的那件事——分割、留样、盲测、把历次接触当作污染源加以隔绝——在艺术学是**病理**：分割原物即毁作品，盲鉴在专业上被视为不负责任。反过来，艺术学每天在做的那件事——知情、读介入层、把历次接触当作证据——在化学是**偏倚**，是方法验证要专门设计实验去排除的东西。一门学问的日课，正是另一门学问诊断为病根的那件事。这一对不能靠判谁对来化解，只能另造一个框架。

教育学与艺术学：教育学要求上一次执行的影响**不影响**结论（所以控制掉）；艺术学要求它**构成**对象的现状（所以登记在案）。前者把它读成噪声，后者把它读成作品的一部分。不能同真。

三家各锁一个终点：**读数、推论链、实践条件**。三个终点互不隶属，谁也不能被折算成谁。而正是在三家争夺终点的时候，它们脚下踩着同一块地，而那块地谁也没有回头看过。

三·三家共同假定的那件事：未涉对照是可以取得的

三家争的是一个判断的效力该由哪一个终点来结算。争得很凶，而争的形式本身暴露了一条三家都没有说出口的假定：

复核所面对的对象，与确立时所面对的对象，是同一个；判断的执行不改变被判断者，或者所改变的那一部分可以被识别并扣除。

把这句话念给三家听，三家都会说“这还用说吗”，而且三家都会立刻指出自己早有装置来处理它——

- 化学说：所以我们留样。
- 教育学说：所以我们用平行卷。
- 艺术学说：所以我们建介入档案。

三套装置，处理的是同一件事，而且三套都依赖同一个东西：一个可以取得的、未受本次判断影响的对照。化学的那一份是另一支等分样，教育学的那一份

是另一套平行题，艺术学的那一份是介入之前的原始层——X光底下那张没被重涂过的底稿。三家的复核之所以能算数，是因为总有一份东西没被这一轮的判断碰过，可以拿来作参照。

本章把这一份叫做**未涉对照**。

到这里，那块共同的地就露出来了：三家争的是终点该设在哪儿，而它们共同假定**未涉对照是可以取得的**；既然可以取得，复核就可以是“沿同一条路径重走一次，到同一个终点”。终点固定、路径可重走、对照可取得，这三条是一件事的三种说法。

现在把这一条搬到中医的临床对象上，它当场断裂：

- **没有等分样**。一个病人不能被分成两份，一份吃药一份不吃，七天后并排比对。
- **没有平行卷**。四诊不是一套可以换一套等值题的题库；换一个医生、换一个时辰、换一个体位，读出来的不是同一个构念的平行形式，是另一次采集。
- **没有介入之前的原始层**。药物的作用不是覆盖在身体表面的一层清漆，可以用溶剂剥掉再看底下；它是进到代谢里去的，舌苔化了就是化了，X光照不出“未服药时的舌苔”。

三家各自的扣除装置，在这里一件也架不起来。而中医的现代验证范式——证候诊断标准的信效度检验、辨证一致性研究、证候疗效评价量表、随机对照与交叉设计——**几乎全部是照化学与教育测量的模子建的**，也就是全部建立在“另取一份”之上。那一份在临床上不存在。

有人会说：那不正说明中医不够科学吗？

不。这正是本章要用第三家的原因：**艺术学证明了，在取不到未涉对照的地方，验证范式仍然可以存在，只是它的形状完全不同**。它放弃盲法，改为逐次登记介入；它放弃“重走一遍”，改为“把历次经过读出来”。一件被修复过七次的祭坛画，今天仍然可以被负责任地归属，靠的不是找到一块没被动过的地方，是靠那七次都有账。

于是本章要处理的问题变得具体了：**在取不到未涉对照的地方，复核的路径该怎么排，以及那个决定复核算不算数的插手点，落在哪一步**。这是一个关于路径的问题，不是一个关于概念的问题。答案不会是“证的本质是什么”，答案会是一条次序，和次序上的一个位置。

四·化学自己的一条材料：没有真值的时候，终点是被这一轮生成出来的

要推翻一条被三家共同踩着的假定，从外面搬一个理由来是没有用的——那只是加入争论，成为第四家，三家可以各自退回自己的阵地。要让三家退不回去，推翻它的那件材料必须是三家之一自己已经掌握、已经发表、只是从来不住这个方向用的事实。

这件材料在分析化学手里，而且是它每天在用的一样东西。

分析化学有一整套评价实验室能力的制度：把同一批物料分发给几十家实验室，各自按自己的方法测，回报结果，再据一个“指定值”给每家算一个偏离分数。这个制度的中枢是那个指定值——它就是“算数”的那个终点。指定值从哪儿来？最好的来路是有证标准物质：一个由更高级别的计量机构定过、带不确定度的值。

但在大量场合没有标准物质。这时国际标准明写着另一条来路：**指定值取自本轮参加者结果的稳健共识**。把这一轮所有实验室的结果放在一起，用抗离群的算法算出一个中心值，那个中心值就是这一轮的“真”。谁离它远，谁的分数就差。

请注意这一条意味着什么。

它意味着在没有独立对照可及的场合，那个“算数”的终点不是固定的，是每一轮由当轮参加者的集合重新生成的。换一批参加者，换一批方法配比，指定值就移动。一家实验室这一轮得分优秀，可能不是因为它更接近某个不动的真值，是因为它这一轮站在人堆的中间。而下一轮若有一批用新方法的实验室加入，人堆的中心会移，它可能原地不动就变成了离群者。

化学人当然知道这一条，而且写进了标准，还配了一整套告诫——共识值在参加者少于某个数时不可靠、方法间存在系统偏倚时要分组算、要报告不确定度。但这一整套告诫回答的是“没有标准物质时怎么公平地评价实验室”。它从来不被拿来回答“复核的终点是不是固定的”。这就是它能推翻共有前提的原因：它在自己的领域里回答的是另一个问题，因此没有人把它读成一条关于验证结构本身的证据。

把它读成一条关于验证结构本身的证据，得到的是这样一句：

当未涉对照不可得时，“算数”的那个终点不是被找到的，是被这一轮的参与结构生成的。

这一句一旦成立，共有前提的三种说法同时垮掉：终点不固定，路径就不是重走，对照也不必假定可得。三家各自的扣除装置随之降级——它们不再是“通向真值的必经之路”，而是在特定条件下（物料可分割／任务可重出／原物有可读的层）才架得起来的三种局部工程。

第二件材料也在手边，来自教育学。教育测量自己发表了几十年的证据说明施测本身改变被测者：做过一次测验的人，在同一构念上的后续表现会系统性地上移；测验不只是测量，它还是一次学习事件。教育学把这一份记为需要用平行卷控制掉的干扰。它没有把它读成“被测对象在上一次判断中已经被改写”的本体证据——因为它有平行卷，它不需要读到那一层。装置的有效性，恰恰是这一层从未被看见的原因。

到这里，本章可以把话说到底了：三家看不见这一层，不是因为它们疏忽，是因为它们各自的核心问题恰好使这一层成为不必要的。化学的核心问题是“这个数准不准”，它有等分样，因此“上次测定改变了样品”这件事对它不是问题——那一份被烧掉了，不影响下一份。教育学的核心问题是“这个解释与使用站不站得住”，它有平行卷，因此“上次施测改变了人”这件事被平掉即可。艺术学最接近，它的核心问题是“这件东西是谁做的、现状是怎么来的”，它已经在读介入层了——但它读的是物上的介入层，而物是不吸收判断的：一幅画不会因为被鉴定为伦勃朗而改变自己的颜料层。只有当被判断者会因为判断而改变自己，这一层才会成为承重的。而临床上的身体，正是这样的对象。

五·并身：一份没有科目的改变

现在可以给那样东西命名了。

一个证被确立之后，它不只是一个记录，它是一道执行指令：据此定治法，据此选方，据此加减，据此叮嘱起居饮食。执行之后，身体变了。这份变化里，有一部分是疾病自己走的，有一部分是这道指令推的。到复核的时刻，两部分混在同一份四诊信息里，没有任何标记可以分开。

复核所面对的那个对象，因此不是初诊那个对象，而是初诊那个对象加上上一版判断的执行产物。而现行的读法，把这个和读成了对象本身。

本章把这份改变叫做**并身**：

并身——上一版判断在它成立的那段时间里，通过它所授权的执行，产生于被判断对象之上、并入被判断对象之内、在复核时被读作对象自身属性的那一份改变。

用会计的话说，它是被**并表**了：子公司并入母公司报表之后不再单独列示，从外面看，报表上只有一个主体。并身的意思正是如此——执行产物并入了对象，从复核那一侧看过去，只有一个身体。

要看清它是一样独立的东西，最快的办法是把它与三个相邻的东西分开：

它不是疗效。疗效是“我想要的那部分变化”，它是有科目的：证候积分下降、主症消失、复发率、有效率，全都有栏位。并身包括疗效，但比疗效大得多——它还包括所有不被当作疗效的变化：味觉的改变、大便性状的改变、睡眠时相的改变、病人对自己身体叙述方式的改变（这一条尤其要紧，因为四诊里有大半是问诊，问诊的输入是病人的叙述，而病人的叙述在一轮治疗后会重新组织）。

它不是不良反应。不良反应也有科目：药物性肝损、过敏、腹泻。凡进入这个科目的，都以“有害”为条件。并身里绝大部分既无害也不有益，只是**改变了下一次读数的基线**。舌苔从厚腻转薄，这不是不良反应，甚至是好事，但它同时意味着下一次辨证再也读不到那个厚腻苔了。

它不是坏病。这一条最要紧，本章下一节专门处理。

并身的关键性质有三条，本章后面全部要用：

1. **它是被正确判断的正确执行所产生的**。不需要辨错，不需要用错药，不需要出任何差错。辨证准确、治法得当、疗效良好，并身照样发生，而且往往更大——**药越对症，身体被改写得越彻底，下一次复核所能读到的未涉部分就越少**。
2. **它不可归咎**。没有任何一方做错事，因此没有任何一方有动机去记录它。一份记录之所以被建立，通常是因为有人需要为某件事负责或免责；并身不涉及责任，于是没有人建这一栏。
3. **它不可撤回**。药已经吃了。没有洗脱期可以把身体还原到七天前——交叉试验里那个“洗脱期”的假定，是药物浓度可以清零，而不是身体状态可以清零。这两件事被长期混为一谈。

三条合起来，构成一个很少见的形状：一样确实发生了、量可能很大、对下一步判断有决定性影响、而且没有任何一方做错事的东西，因此它在所有账本上都不设字段。

六·为什么有过错的那一份有科目，无过错的那一份没有

中医对这件事并非全无所知。恰恰相反，两千年前它就有个正式的科目，而且位置极高。

《伤寒论》说：太阳病三日，已发汗，若吐、若下、若温针，仍不解者，**此为坏病**。紧接着给出的处置是那十二个字——**观其脉证，知犯何逆，随证治之**。

把这一段读慢一点，会发现它是一套完整的复核程序，而且是本章所需要的那一套：

- “已发汗，若吐、若下、若温针”——先清点上一版执行做了什么。
- “仍不解者，此为坏病”——认定现在这个状态已不是原来那个病，它是被**执行改写过的**。
- “知犯何逆”——把执行产物单独立一笔账，问清楚是哪一次动作造成了哪一处偏移。
- “随证治之”——在扣除之后，对剩下的部分重新判断。

这是一条与本章所要论证的路径完全同构的次序：清点执行 → 认定对象已被改写 → 单独结算执行产物 → 再判断。它比本章早了一千八百年。

那么本章还剩什么可说的？

剩的正是那一处分界，而这一处分界是整篇文章的枢纽：

坏病是“误”的产物；并身是“正”的产物。

坏病成立的条件是“犯逆”——该汗而下、该下而汗、当温反清。它是一个**可归咎**的科目：有一次错误的动作，有一个责任人，有一条可以追溯的因果。正因为可归咎，它才有资格进入记录——中医的整个“误治”传统，从《伤寒论》的坏病到后世的医案医话，记的都是这一类。

而本章说的那一份，发生在**没有犯逆**的情形下。辨证对，治法对，方对，剂量对，服法对，病人依从性满分，七天后主症显著缓解——**在这个最理想的场景里，并身达到最大值**。因为最有效的干预正是改写最彻底的干预。

于是有了一个结构性的空缺：

	上一版执行的产物	有无科目
有过错	坏病、误治、药邪、不良反应、医疗差错	有（可归咎，责任驱动记录）
无过错	本章所说的并肩	无（不可归咎，无人需要为它负责）

记录制度是被责任驱动的，因此它只覆盖可归咎的那一半。无过错的那一半量更大、更常见、对复核的干扰更系统，却因为没有人需要为它免责或追责，从来没有人为它开一栏。

这就是并肩之所以“不算一样东西”的机制：它不是被记得少，是被拆散记进了别的两个科目里——好的那部分记进“疗效”，坏的那部分记进“不良反应”，剩下大量既不好也不坏、只是改变了下一次读数基线的那部分，哪里也不记。删掉本章提出的这个名字，它不会变得难测，它会重新变成三样彼此无关的东西：一点疗效、一点副作用、一片没人看的空白。

同样的空缺可以在三门源学科里逐一找到，位置完全一样：

- **化学**：被离群检验剔掉的那一次测定，只在“作废”这个意义上被记录——它作为一次**曾经被当真过的读数**、以及它可能已经指导过的下一步操作，不留一笔。
- **教育测量**：评分者之间的分歧被折算成误差方差，合意之后分歧本身不留字段；评分者培训中被“训掉”的那些判断，没有任何档案保存它们曾经存在。
- **艺术学**：一幅从某位大师名下被移走的画，那次移动之前它“曾经算数过”的那几十年——挂在哪个展厅、进过哪本图录、影响过哪些研究——在新目录里不占一行。

三处空缺是同一个形状：**已经生效过、已经推动过下一步动作、随后被判定为不再有效的那一版判断**，它在生效期间的产出没有科目。中医的特殊性只在于，那个产出直接落在被判断者的身体上，因而它必然回到下一次判断的输入里。

七·两根轴与四格

把上面的东西收成一张辨别表，需要两根轴。

第一根轴：未涉对照可不可得。 复核所要依据的证据里，有没有一部分是在上一版判断执行之前就已固定下来的、或者结构上不受其影响的。

第二根轴：上一版执行的产物有没有被单独立账。 是“另有一笔记着它”，还是“并入对象、不单独列示”。

两根轴切出四格：

	单独立账	并入对象
未涉对照可得	① 可拆分复核 ——既有未被动过的对照，又有上一次执行的完整记录。化学的留样复测在这一格：留样在冰柜里，上一次的谱图在系统里。	② 闲置的对照 ——底留下了，但复核程序不去读它。临床上最常见：初诊拍了舌照、录了脉图，复诊时没有任何流程要求调出来比对。
未涉对照不可得	③ 知情复核 ——取不到未被动过的对照，改为把历次介入逐次登记，在知情的前提下作判断。艺术学的保存修复档案在这一格；《伤寒论》“知犯何逆”也在这一格。	④ 靶格

第④格是本章要打的那一格：**取不到未涉对照，而且上一版执行的产物被读作对象自身的属性。**

八·靶格：全部形式审查都能通过的那一格

第④格不是一个坏的格子。它是一个看上去最好的格子。这是它值得被单独写一篇文章的全部理由。

一个落在第④格的诊次，形式审查全部通过：

- 辨证有依据，四诊记录完整；
- 治法与证相符，方药与治法相符；
- 病人依从性良好，七剂服完，无漏服；
- 复诊按时，四诊重新采集，记录规范；
- 复核结论自治，与新的四诊信息完全吻合；

·若有质控，医生间一致性达标——因为不同医生读的是同一份复诊四诊资料，他们会读出同一个证。

没有任何一个环节会报错。而恰恰在这样一个诊次里，复核所依据的证据里可能**一条都不是**在首次处方之前固定下来的。

它有三条签名，逐条对照：

签名一：它在短期指标上表现为改善。疗效越好，并身越大。一位全程管理规范、依从性满分、症状缓解显著的病人，他的每一次记录都在治疗之中——因为规范化的全程管理意味着不留空窗。相反，一位中途自行停药两周才来复诊的病人，那两周反而提供了一段可用的、接近未涉的观察。**越是模范病例，可用的未涉对照越少。**

签名二：它的失效不产生任何信号。如果第④格的复核给出了一个错误的方向，谁会知道？没有人。因为新的证与新的四诊完全吻合，新的方与新的证完全吻合，病人的反应又会成为下一轮的四诊。**这是一个可以自治地走很多轮的闭环，而闭环内部不产生任何异常信号。**唯一会暴露它的场合，是病人换了一位医生、而那位医生恰好问了“你上一版吃的是什么”——但那时通常已经过了很多轮。

签名三：它自我加固，而且越正规化越严重。这一条最要命。复核的可信度在现行方法论里靠什么保证？靠**独立性**：复核者不应被上一版判断影响，最好不知道上一版判出的是什么证，以免锚定。这是所有验证范式的共识，而且它在别处确实是对的。

但在这里，独立性的要求与可归因性的要求正面冲突：

复核要有效，它必须独立于上一版判断；复核要正确，它必须知道上一版执行过什么。

盲，则读不出并身；不盲，则被锚定。三门源学科各有各的解法：化学与教育学用“另取一份”同时满足两边——复核者可以完全盲，因为他手上那一份本来就没被动过；艺术学**放弃盲**——它承认原物只有一件，因此把知情作为前提，靠逐层登记来抵抗锚定。中医现行的方法论，学的是前两家，而临床对象的结构属于后一家。

于是产生了一个精确的失配：中医的复核越是向着“**更独立、更盲、更标准化**”的方向被改进，第④格就越牢固。每一次强调“复核者不应查阅上一版辨证

结论”，都在切断读出并身的唯一通道。这不是执行不力，这是方向本身在把问题做大，而且每一步都能通过审查。

本章要打的正是这一格。而打它的方式不是加强复核，是在复核之前动手。

九·留底率：一个三门学科量纲不同而比率相同的读数

一个判断若停在“有这么一样东西”，它还不能被检验。要能被检验，需要一个可以事后清点的数。

本章给出的读数是留底率：

留底率 = 一次复核所依据的证据项中，取自“上一版判断执行之前已经固定下来、或结构上不受其影响”的那部分项数 ÷ 该次复核所依据的证据项总数。

“留底”取的是两重意思：档案上的留底存查，与化学上的留样。两者是同一个动作——**在动手之前，先把一份东西固定下来。**

它在四个领域里的清点方式，量纲各不相同，比率相同：

- **分析化学**：一次复核性测定所用的样品中，取自第一次测定之前分装、且未经处理的留样，占比多少。
- **教育测量**：一次复核所依据的表现证据中，取自干预开始之前的基线、或取自与干预内容无重叠的独立任务，占比多少。
- **艺术学**：一次技术复核所读的层中，属于历次介入之前的原始层（底稿、原始画布、初始颜料层）的，占比多少。
- **中医临床**：一次复核所依据的四诊与检查信息中，属于首次处方之前已记录、或结构上不受该处方影响（既往病史、体质辨识、既往影像与化验、家族史、发病诱因叙述）的，占比多少。

四处的单位分别是“份”“任务”“层”“条”，量纲毫不相干；而比率是同一个东西：**这一次判断，有多少是站在未被自己动过的地方作出的。**

一个新读数要能立住，必须过四关，逐一交待：

第一关，无量纲。是比率，四个领域可以并排比较。一位医生的留底率 0.15，一间实验室的留底率 0.60，可以放在同一张图上。

第二关，可事后清点。这是最要紧的一关，也是本章特意把定义写得笨拙的原因。它不需要任何新仪器、不需要任何前瞻设计，只需要两样已经存在的东西：**病历上每一条记录的书写时点，与处方的开具时点。**拿一份病历，把复诊那一栏所依据的每一条信息列出来，逐条问“这一条是什么时候写下的”，早于首方者计入分子。多数电子病历系统的时间戳已经存在，不必新增字段。

第三关，把一个印象变成一个数。“这次复诊其实看的是药后的身体”是一句印象，人人有感觉，无人能拿它做任何事。0.08 是一个数，可以排序，可以比较，可以进研究设计，可以在两家医院之间对照。

第四关，与既有主要指标正交。这一关最容易漏，也最容易毁掉一个新指标——如果它与某个现成指标高度相关，它会立刻被读成那个指标的一个代理，白造。

留底率过这一关，而且过得很干脆，因为它与所有既有质量指标呈负相关：

一位病人，疗效评价满分，症状积分从 24 降到 3，随访完整率 100%，依从性 100%，复诊七次一次不落，病历书写质量评为优秀——**他的留底率可以是 0.04。**因为初诊只有一次记录，此后每一次四诊都在服药之中；病历写得越勤、随访越密，分母越大而分子不变。

>

另一位病人，中途自行停药十八天，随访脱落一次，依从性评为差——**他的留底率是 0.31。**因为那十八天里的一次急诊记录，是在药物作用基本消退之后写下的。

依从性最好、管理最规范的病例，留底率最低；管理最“差”的病例，反而留下了可用的未涉观察。这一条足以说明它不是任何现有指标的影子。它也预告了一个必须正面处理的伦理风险，见第十七节。

留底率不是越高越好，这一点必须写死。它是一个**结构读数**，不是一个绩效指标：它说明的是“这一次复核站在多硬的地上”，而不是“这位医生做得多好”。一个急症、一个必须立刻用药的场合，留底率天然接近零，而那是完全正确的临床决策。**把留底率当成考核分，正是本章最担心的那种后果。**

十·一句不含情态词的问话

一个判断如果只能靠“应当”“充分”“真正”“恰当”这类词来表述它的判据，那它还没有落到可查的层面。本章的判据是一句可以直接拿去问病历、问日志、问流程文件的问话，里面一个情态词也没有：

这一次复核所依据的记录里，有哪几条是在上一次开方之前就已经写下的？

它是一句问事实的话，不是问判断的话。回答它不需要任何人的意见，只需要翻记录。

把它在五个场合各跑一遍，答案互不相同——这一点很重要，因为如果五个场合答出同一个东西，那说明它只是承重命题的复述，不是一个能分辨情况的判据：

场合一，中医复诊。答案通常是：既往病史与体质辨识两条，其余全部写于服药之中。**留底率约 0.1 或更低。**值得注意的是，这个答案在门诊与住院之间差异极大——住院病历因为有入院当日的系统查体，留底通常厚得多。

场合二，实验室复测。答案是：留样一支，第一次测定之前分装、冷冻保存。**留底率接近 1。**但若这份物料是易变基质、留样期间发生了降解，那份留样也不再是未涉的——它没被本次判断动过，却被时间动过。**这暴露出未涉对照有两种失效方式：被自己动过，与被时间动过。**这一条是查出来的，不是从命题推出来的。

场合三，学期末对一位学生的复评。答案是：入学摸底测验一次、学期初的一次作品。而这两样通常已经被教师翻看过、评讲过、据此调整过教学——**未涉对照存在，但它已经被读过，因而它对教师的判断不再独立。**这是第三种失效方式：对照在物理上未被动过，在认知上已经被动过。这一条也是查出来的。

场合四，一幅经过修复的画的重新归属。答案是：X 光与红外能读出的底层若干处，加上一份一九五三年的修复报告。**留底率可以远高于零，而且它是靠仪器“挖”出来的，不是靠当初留下的。**这提示了一条本章原本不会想到的路：**未涉对照有时可以事后重建，只要对象上留有分层。**身体上有没有类似的分层？有一些——毛发的重金属分段、糖化血红蛋白的时间窗、影像的历史序列。这一条同样是查出来的。

场合五，一次生产环境的故障复盘。答案往往是：没有。因为第一批日志在热修复过程中被轮转覆盖，而故障现场在修复动作之后已不可复现。**留底率为**

零，而复盘报告照样写得出，且写得完全自治。这是第④格在另一个领域的原样重现。

五个答案互不相同，其中至少三个（时间致失效、认知致失效、事后重建）不能从承重命题直接推出——它们是跑出来的。

十一·复核路径的重排：可插手的那一步落在哪里

到这里可以正面回答那个问题了：一个证如何被确立并被复核，才算数？

现行的默认路径是两步：

确立 → 复核

本章主张它实际上是四步，而中间两步现在没有位置：

确立 → 执行 → 并身结算 → 复核

四步各自的动作：

1. **确立**：四诊合参，判出证，形成一道执行指令。
2. **执行**：治法落地，身体被改写。这一步在现行记录里只留下“用了什么方”，不留下“改写了什么”。
3. **并身结算**：在复核之前，先把上一版执行的产物单独清点一遍——哪几条现有表现是这道指令的直接产物，哪几条与它无关。这一步现在**完全没有位置**。
4. **复核**：在扣除已结算部分之后，对剩余部分重新判断。

三个要点：

第一，可插手的那一步不是第四步，是第三步。现行的全部改进努力都压在第四步上——提高辨证一致性、统一证候诊断标准、加强复核者培训、强调独立与盲法。这些改进各自都对，但它们全部作用在一个已经被污染的输入上。**输入没有被拆分，加工得再精细也只是把并身更精确地读成对象。**

第二，第三步不是靠回忆做的，是靠第一步之前留下的东西做的。这是本章与《伤寒论》“知犯何逆”的唯一实质差别，也是本章对那十二个字的补充。“知犯何逆”是一个**回溯**动作：事后想清楚上一步做了什么。回溯在误治的场合可行，因为误治有明显的转折点（一发汗，人就变了）；在正治的场合不可行，因为正治的改写是渐进的、符合预期的、与病程自愈方向同向的，回忆分不

开。能把正治的改写分出来的，只有一样东西：动手之前留下的那一份。所以第三步的可行性，实际上由第一步之前的动作决定。

第三，因此真正的插手点在时间上比复核早得多。用一句话说：

复核的可靠性，不是在复核那一刻挣来的，是在第一次开方之前留下来的。

这句话把整个问题的重心搬了家。它意味着“如何复核”这个问题的答案，大部分不在复核里。它也解释了为什么加强复核的努力长期收效甚微——**它们都在一条已经过了岔路口的路上加速。**

三门源学科的路径在这个框架下可以被重新排列，各自的有效性来源也清楚了：

- **化学**走的是“先固定条件，再固定操作，最后看落点”；它之所以能把第三步省掉，是因为它在第一步之前就分了样——**它的留底动作藏在取样制备里，早得让人看不见。**
- **教育测量**走的是“先看表现，再看情境，最后检查推论链”；它之所以能省掉第三步，是因为平行卷把“另一份”造了出来——**它的留底动作藏在题库建设里。**
- **艺术学**走的是“先看物，再看流转，最后落回实践条件”；它省不掉第三步，所以它把第三步做成了一整个专业——**保存修复档案就是并肩结算的制度化形式。**

三家都有留底动作，只是两家藏得太深，深到学习者只学到了后面几步。中医现代化学的正是那后面几步。

十二·六个可以立刻做的记账动作

一条判断如果给不出低成本的落地动作，它就只是一个说法。以下六条按成本从低到高排列，前三条不需要任何新增设备与新增流程。

一、**给每一条记录标一个“相对首方的时点”**。不是新增字段，是把已有的时间戳与处方时间做一次减法：这条记录是首方前第3天，还是首方后第8天。仅此一项，留底率就可以被自动算出。

二、**复诊记录里固定一栏问诊：上一版的药服了几剂、什么时候停的、停后到今天几天。**这句话每位有经验的医生本来就在问，只是答案不落在结构化字段里。落进去，“洗脱窗”才可能被识别。

三、初诊时把最容易变的三样东西留下一份可复看的记录：舌象照片、脉图或脉诊描述的固定格式、主症的原话摘录。留原话尤其重要——问诊的输入是病人的叙述，而叙述在一轮治疗后会被重新组织，原话是唯一能留住这一层的东西。

四、在病历上区分两种复诊：药中复诊与药后复诊。前者在药物作用期内，后者在停药一定间隔之后。二者的四诊信息不应被当作同一类证据使用。这一区分不需要新技术，只需要一条书写规范。

五、对拟长期用药者，设计可容许的“观察窗”。在临床安全允许的前提下，安排一次不在药物作用期内的四诊采集。这一条必须极其谨慎，因为它直接触及临床伦理——任何以提高留底率为目的的停药都是不可接受的（第十七节）。可行的做法是利用已有的自然窗口：换方之间的间隔、节假日中断、剂型转换期。

六、把“上一版判断”写进复核的输入，而不是把它藏起来。这是与主流方法论相反的一条，因此需要配套：复核者应当知道上一版判的是什么证、用的是何方，但要求他先书面写出“若上一版执行有效，我预期今天会看到什么”，再采集四诊。写在前面的预期，是抵抗锚定的手段——它把“知情”从一个偏倚源，改造成一个可被检验的预测。这条做法在别处已有成熟形式，本章第十三节会交代它的来路并划清界线。

六条里，第一条与第六条最便宜也最要紧：一条让并身可以被清点，一条让并身可以被读出。

十三·与最近的几种说法的分界

一个新名字最常见的下场，是被读者正确地认出来是某个旧概念的改名。要避免这个下场，只有一个办法：把最容易与它混淆的说法逐个请进来，各自说清它到哪一步，再说清同一个案例上两边会给出什么不同的判断。

在做这件事之前，先把本章的命题剥掉中医的一切名词，只留下形式：

一个判断的执行改变了它自己的检验对象，而检验程序把这份改变读作对象本来的样子；且检验程序越是遵守独立性规范，这份改变越读不出来。

这个纯形式的结构，在若干门上游学科里各有名字。逐个交待。

(一) **交叉试验的残留效应与洗脱期** (临床试验方法学, Grizzle 1965; Senn 2002 *Cross-over Trials in Clinical Research*)

这是本章最近的方法学邻居, 也是最容易吞掉本章的一位。交叉设计让同一受试者先后接受两种处理, 方法学早就承认前一期处理可能残留到后一期, 处理办法是插入洗脱期, 并在模型中检验处理 $\times$ 期次交互项。

它到哪一步: 它承认残留存在, 并把残留当作一个可以被时间清除、可以被统计模型估计的量。它的整套装置建立在“药物浓度归零则状态归零”之上。

分离线: 本章说的这一份不随浓度归零。舌苔转薄之后不会因为停药三十天而重新变厚成原来那个苔; 病人对自己身体的叙述方式一旦重组, 不会因为停药而复原。**残留效应假定对象有一个可回归的基线, 本章说的是基线本身被移动了。**另一处差别更实际: 残留效应是研究设计里的一个项, 它只在有平行期次可比时才被估计; 临床复诊没有期次, 没有随机化, 没有另一臂。

判定形式: 取一批停药后间隔不同天数复诊的病例。若“复核结论相对上一版的偏移量”随停药间隔延长而单调回落至零, 则残留效应对而本判断错; 若偏移量在停药间隔延长后趋于一个不为零的平台, 且该平台高度与上一版治法的改写强度呈剂量-反应, 则本判断对。

(二) **测试效应与练习效应** (教育心理学, Roediger & Karpicke 2006)

它到哪一步: 它证明施测本身改变被测者, 且量级可观。它的处置是用平行卷、等值题库把这一份控制掉。

分离线: 它处理的是**测量动作**的改变作用, 本章处理的是**执行动作**的改变作用。测验改变人是测量的副产品, 医生开方改变人是判断的**目的**。二者的方向相反: 前者要设法减小, 后者不能减小——减小它就等于不治病。一个**不能被减小的干扰项, 不能用控制它的方法处理, 只能用记账的方法处理。**

判定形式: 若在完全不施加任何治疗、仅重复四诊采集的对照组中, 复核结论的偏移量与治疗组相当, 则本判断被测试效应吸收; 若治疗组的偏移量显著大于纯采集组, 且差值随治法强度递增, 则本判断对。

(三) **治疗前活检原则与治疗掩蔽** (临床医学)

血液科有一条硬规矩: 怀疑淋巴瘤者, 激素在活检之前不得使用, 因为激素会使病理形态无法诊断。感染科同理: 经验性抗生素之后血培养转阴, 病原不再可及。

它到哪一步：这是医学自己最接近本章的一位，而且它已经把“执行毁掉证据”这件事做成了操作规范。

分离线：它只覆盖**证据被抹掉**的情形——原来能读出的读不出了，而且**读不出这件事本身有信号**（病理科会报“形态不典型，建议结合治疗史”）。本章覆盖的是**证据被改写**的情形——依然能读出一个完整、自洽、可判读的结果，只是这个结果的一部分是上一版执行的产物，而且**没有任何环节会报出异常**。掩蔽是一个有信号的失效，并身是一个无信号的失效。这也是为什么掩蔽有规范而并身没有：**有信号的失效才会被制度处理**。

判定形式：若一个诊次的复核异常总是伴随某种“信息不可读”的提示（记录缺失、指标无法判读），则本判断被治疗掩蔽覆盖；若在全部记录完整、全部指标可判读、无任何提示的诊次中，复核结论的偏移仍与执行强度相关，则本判断对。

（四）医源性损害（Illich 1975 *Medical Nemesis*）

它到哪一步：它指出医疗行为本身制造疾病，并把这一现象扩展到社会与文化层面。

分离线：医源性损害以**有害**为构成条件；本章所说的那一份以**无过错、常常有益**为典型情形。落在“有害”上的部分已经有科目（不良反应、医疗差错），恰恰不是本章要立的那一栏。**本章要立的是无害那一栏，而它之所以没有栏，正是因为它无害**。

判定形式：同一份病例，按医源性损害的框架判定为“无医源性问题”（疗效好、无不良反应），而按本判断判定为“高并身、低留底率、复核依据不足”。两边对同一份材料给出相反的评价——若临床上找不到这样的病例，本判断被覆盖。

（五）循环效应（科学哲学，Hacking 1995/1999, *looping effects of human kinds*）

它到哪一步：它指出人类分类与被分类者之间存在互动回路——被归入某一类的人会因为知道自己被这样归类而改变，分类因此需要修订。这是本章形式层上最著名的一位。

分离线：循环效应的两个端点是**分类与人**，中介是**知晓**——人知道了自己被怎样归类，因而改变。本章的两个端点是**执行与对象**，中介是**药理与生理**，与知晓无关：一位对自己被判为何证毫不知情、甚至语言不通的病人，同样发生并

身。更要紧的一处：循环效应描述的是一个**互动**，因而它预设两端可分；本章说的正是**两端不可分**——执行产物并入对象之后，“分类”与“被分类者”之间不再有两个可区分的端点。

判定形式：若并身量随病人对自身诊断的知情程度递增（知道自己被判为某证的人偏移更大），则循环效应对；若在完全不知情、无语言交流的病例（如婴幼儿、意识障碍者）中，并身量与知情病例相当，则本判断对。**这是一条可以直接做的对照。**

（六）自证预言（社会学，Merton 1948）

它到哪一步：一个起初为假的判断，通过它引发的行为使自己变成真的。

分离线：自证预言要求判断**变成真的**；并身不要求。一个被推翻的证，同样在成立期间产生了并身——**它的产物留下了，而它本身被否定了**。自证预言处理的是判断与事实的符合关系，本章处理的是判断的产物在下一次判断中的账目位置。二者甚至可以同时反向发生：一个自我否证的判断（用药之后该证的表现消失了），其并身量最大。

判定形式：若并身总是使复核结论更接近上一版判断，则自证预言对；若在大量病例中并身使复核结论**远离**上一版判断（且这正是“效不更方”与“效则更方”之争的实质），则本判断对。

（七）述行性 / 操演性（经济社会学，Callon 1998；MacKenzie 2006 *An Engine, Not a Camera*）

它到哪一步：它指出理论不只是描述市场，还参与建构市场——期权定价模型被采用之后，市场价格向模型靠拢。这是形式层上最有力的一位外圈邻居。

分离线：述行性的机制是**行动者按照模型行事**，因而它要求模型被广泛采用、被内化、被制度化；本章的机制是**一次具体的执行落在一个具体的对象上**，不需要任何采用与传播，一位医生对一位病人就完全成立。规模条件完全不同。另一处：述行性关心模型的真假被改变，本章关心检验的输入被污染。

判定形式：若并身量随该证型在共同体中被采用的广度递增，则述行性对；若在使用极冷门证型、共同体几乎不用的个案中，并身量与常用证型相当，则本判断对。

（八）可审计性与审计社会（会计学与组织社会学，Power 1997 *The Audit Society*）

这一位必须点名，因为本章的形式剥出来是“越规范越读不出”，属于“越成功越失败”这一类反转结构，而这一类结构在管理学与组织社会学里早有若干成熟的名字——成功陷阱(Levinthal & March 1993)、能力刚性(Leonard-Barton 1992)——其中最贴近本章的是 Power 的可审计性论：审计不是中性的检查，它反过来改造对象，使对象变得“可被审计”。

它到哪一步：它说明检验制度会重塑被检验者，被检验者为了通过检验而调整自己的形态与记录。

分离线：方向相反。可审计性的机制是**被检验者主动迎合检验**；本章的机制是**被检验者被动承接执行的产物，而且无人迎合**——病人没有为了通过复核而改变自己，医生也没有为了让复核好看而做任何事。更要紧的一处：可审计性的产物是**为检验而生的表象**，若掀开表象，底下那个东西还在；并身的产物是**真实的生理改变**，没有表象可掀。

判定形式：若观察到被复核方为通过复核而调整记录、措辞或表述（迎合痕迹），则可审计性对；若在完全无迎合动机的场合（如回顾性调阅十年前的病历、当事人已不在岗），那份改变依然使复核依据不足，则本判断对。

（九）分期迁移与威尔·罗杰斯现象（流行病学，Feinstein, Sosin & Wells 1985）

它到哪一步：检查手段变精细之后，一部分病例从低分期移入高分期，使得两组的统计表现同时改善，而没有一个病人变好。它是“读数变化不等于对象变化”的经典案例。

分离线：分期迁移的原因在**判据侧**——分期标准或检测手段变了，对象没动。本章说的原因在**对象侧**——判据没变，对象被我们自己推动了。二者可以同时发生，而且方向可以相反。

判定形式：固定同一版证候诊断标准、同一套采集方法，若复核结论仍系统性偏移，且偏移量与执行强度呈剂量-反应，则本判断对而分期迁移不足以解释；若在完全无治疗介入的纯观察队列中出现同等偏移，则应归于判据侧。

（十）先做预测再看反馈（认知心理学，Kornell, Hays & Bjork 2009；errorful generation / pretesting effect）

这一位不是本章命题的邻居，是本章第十二节第六条做法的**来路**，必须交待清楚，否则那条做法会被读成本章的发明。

它到哪一步：先让人写下自己的判断或猜测，再给反馈，学习效果优于直接给答案；写在前面的判断能抵抗后来的锚定。

分离线：它是一条关于**学习**的规律，本章借它做的是一件**记账**的事——让复核者先写下预期，目的不是让他学得更好，是把“知情”从偏倚源改造成可检验的预测，从而使知情复核在方法上成为可能。**本章对它的关系是借用，不是竞争。**明写这一点，是为了不把一条几十年的成熟结论算进自己账上。

十三之一·划完之后：这些邻居各自看不见什么

把上面每一位再问一遍“它把什么当作给定，因此它结构性地看不见什么”，句式固定：

- **交叉试验方法学**把“存在一个可回归的基线”当作给定，因此看不见“基线本身被移动”——一旦承认基线不可回归，洗脱期这个装置就没有意义，整个交叉设计的适用条件要重写。
- **教育测量**把“任务可重出”当作给定，因此看不见“没有平行形式的判断如何复核”——一旦承认存在大量不可重出的对象，效度论证的概化那一环就失去抽样框。
- **能力验证制度**把“参加者集合是随机的一批人”当作给定，因此看不见“终点由参与结构生成”这件事的后果——一旦承认这一点，跨轮次的能力比较就不再有意义。
- **循环效应**把“分类与被分类者是两个可区分的端点”当作给定，因此看不见“两端已经合并”的情形——一旦承认合并，它所描述的那个回路就没有两个节点可以连。
- **可审计性论**把“被检验者有迎合的动机与能力”当作给定，因此看不见“无人迎合而检验依然失效”的情形——一旦承认这种情形，它对失效的全部解释（制度压力、表象生产）就落空了。
- **医源性损害**把“值得记录的改变以有害为条件”当作给定，因此看不见无害的改变——一旦承认无害的改变也必须记账，它的整个批判框架就失去了道德杠杆。

六行背后是不是同一块地？是。它可以被写成一句：

凡是被验证的对象，在验证程序之外有一个独立于验证而存在的状态；验证只是去看它，看的动作不构成它。

这句话没有任何一派会去质疑，因为互相批判的各派都站在它上面：主张更严格盲法的与主张知情鉴定的，主张统计控制的与主张档案登记的，全部预设了那个“独立存在的状态”是有的，分歧只在怎样更准确地接近它。“未被任何学科辨识”的真正来源不是没有人看，是看的人都站在它上面。

本章所说的那一份，正是从这块地的缝里掉下去的：它既不属于对象自身，也不属于验证程序，它是执行的产物——而执行既不在“对象”这一栏，也不在“验证”这一栏。

十四·与本系列各章的分界

本系列近期有数篇用同一组学科、同一路数处理相邻问题，它们是本章最近的邻居，必须逐篇划清，否则读者有理由怀疑这些文章在换着说法说同一件事。

《撑平》（《治未病立账学》第一章）（同组学科，处理“何谓未病”）。那一篇说的是：一个看起来正常的读数，可能由持续的双向抵消维持着，而维持这份持平的工作在账本上不设字段；它给出的读数是持率。**分界**：他持率问的是“当下这一态由谁在维持”，留底率问的是“这一次判断依据的证据由谁产生”。两个量正交：一位撤除支持即恶化的病人（他持率接近1）可以有很高的留底率，只要他的基线记录完整；一位已经痊愈、完全不需要任何支持的病人（他持率为0）可以留底率为零。**判决性对照**：在一组他持率一致的病例中，若留底率对复核结论偏移量没有独立解释力，则本章被那一篇吸收。

《代持的病》（《治未病立账学》第四章）（同组学科，处理“未病如何被识别、无症状时如何正当干预”）。那一篇管的是**识别侧**——一个尚未显现的东西如何被辨认出来并被正当地处理；本章管的是**复核侧**——一个已经被辨认、已经被处理过的东西，如何被第二次判断。两篇在时间轴上首尾相接：那一篇结束于“作出判断并干预”，本章开始于“干预已经发生”。**判决性对照**：那一篇预测识别的困难随症状显现而下降；本章预测复核的困难随干预有效性而上升。同一位病人，干预越成功，那一篇越乐观，本章越警惕。这是两篇对同一份材料给出相反读数的地方。

《无事对价》（《治未病立账学》第七章）（同组学科，处理“治未病为何两千年最高纲领却实践得最少”）。那一篇说的是：预防兑换的是非事件，而账本只能记事件，因此一次成功的预防不产生下一轮的资源。**分界**：那一篇的缺栏在**回报侧**（没有发生的事件没有科目），本章的缺栏在**对象侧**（已经发生的改变

没有科目)。两个量正交且可同时为极值：一次成功的预防，无事对价很高而留底率不受影响——因为没有执行任何改写身体的干预；一次成功的治疗，无事对价为零而留底率降到接近零。**判决性对照**：若在纯预防性干预（未用药、仅生活方式指导）中留底率同样显著下降，则本章的机制被那一篇的机制覆盖。

《**阈前距**》（《**治未病立账学**》第二章）（另一组学科，同一问题域）。那一篇处理的是判据线本身在移动、两端都在动的情形。**分界**：线移发生在**判据侧**，并身发生在**对象侧**。**判决性对照**：固定同一版诊断标准、同一套采集规范（即锁死判据侧），若复核结论仍系统性偏移且偏移量与执行强度相关，则本章对而线移不足以解释此处的现象。

《**临界回落**》（**本系列单篇**）（教育学×化学×管理学，处理“为何课堂上真正的学习很少发生”）。那一篇处理的是抵达了临界而没有被接住，因此回落。**分界**：那一篇的对象是**没有被承接的到达**，本章的对象是**已经被执行的判断**。两者互补而非重叠：那一篇的缺失发生在过程完成之前，本章的缺失发生在过程完成之后。

把五篇并排看，它们共有一个形式——**发生了而没有产生条目的东西**——而各自落在不同的位置：维持工作、判据线的移动、代持额度、无事对价、并身。共有形式不构成重复，位置不同才是要害；但共有形式也意味着，若哪一天有人证明“缺栏”这个说法本身是一个万能钥匙、放在任何领域都开得动，那么这一整批文章都要一起重估。本章把这句话写在这里，是因为它是这一批文章共同的、最危险的证伪条件。

十五·三门学科反过来给本章加的约束

借用不是单向的。三门学科在被本章使用的同时，也各自砍掉了本章原本打算主张的一些东西。三处削减都写在这里，并写清削掉的是什么。

第一处，艺术学削掉了本章最自然的那个处方。

本章原本会写下这样一句更强的话：“**只要把上一版执行的产物单独立账，复核就可以把它拆出来。**”这句话很顺，也很像一个可以直接推行的政策。

艺术学不允许它成立。艺术学有一整套介入档案制度，逐次登记每一次修复：哪一年、谁、什么溶剂、动了哪一块。档案完备到今天仍在用。而它的实际经验是：**登记介入并不能防止介入被读进作品**。十九世纪对一批祭坛画的过度

清洗有完整档案，今天的观众看到的仍然是被清洗之后的那件东西；学界对“该不该恢复到清洗之前”永远无解，因为清洗之前那一层已经不存在了，只剩下关于它的记载。**档案保住的是可追问性，不是可拆分性。**

因此本章的主张必须削弱为：**立账保证的是这份改变可以被追问、可以被纳入判断，不保证它可以被扣除。**这是一个实质性的削弱：它意味着即使第十二节的六条全部落实，复核也不会变成“重走一遍”，只会变成“知情之下的重新判断”。留底率因此不是一个可以被做到1的指标，是一个说明“这次判断站在多硬的地上的读数。

第二处，化学动了本章的适用范围与价值排序。

本章原本会写下这样一句更强的话：“一切验证都建立在一个不可得的对照上。”这句话有气势，而且很容易从本章的论证滑出去。

化学不允许它成立。化学的等分样范式在可分割物料上完全有效，而且有效了一百多年：一份矿石可以被破碎、混匀、缩分成互不影响的若干份，每一份都真的是未涉对照。**在那里，未涉对照不是不可得的，是被制造出来的。**化学甚至有一整套关于“怎样制造代表性”的理论。

因此本章的适用范围必须收窄为：**仅当被判断的对象不可分割、不可重出，且判断的执行作用于对象本身时。**落在这个范围之外的，本章的命题不适用，也不应被援引。同时，价值排序也被动了：本章不主张“留底率越高越好”这一普遍原则——在可分割的场合，追求留底率是白费力气，因为造一份新的比留一份旧的更可靠。

第三处，教育学把本章从一个二值命题改造成一个连续量。

本章原本会写下这样一句更强的话：“未涉对照在临床上不存在。”

教育学不允许它成立。练习效应的量级研究说明，平行卷可以把大部分而非全部的测试效应控制掉——也就是说，“未受影响”从来不是一个是非题，是一个程度问题。同样地，临床上“结构上不受首方影响”的证据并非零：既往史、体质、家族史、发病诱因的原始叙述、首方之前的影像与化验，都在不同程度上属于这一类。

因此本章必须放弃“不可得”的绝对表述，改为一个连续的比率。**这一处削减改变的是命题的形式本身：**它使本章从一条“临床复核不可能可靠”的悲观断言，变成一条“复核的可靠性有一个可以被清点、可以被改善的下限”的可操作命题。三处削减里，这一处最伤气势，也最有用。

十六·不适用的边界

本章的命题在下列情形不适用，援引者应当自己先排除：

一、**对象可分割或可重出的场合**。实验室检测、批量物料、可重复施测的标准化任务，请直接用各自成熟的范式，本章不提供任何增量。

二、**急症与危重症**。这类场合留底率天然接近零，而且理应如此——先救命，不留底。在这里追求留底率是错误的，甚至是危险的。本章的适用场景是慢病管理、长期调理、反复复诊这一类允许时间的场合。

三、**单次判断、不涉及复核的场合**。没有第二次判断，就没有本章所说的

问题。

四、**执行不作用于被判断对象的场合**。若一次判断的执行落在别处（例如一次行政评估的结论只影响资源分配，不影响被评估者本身），则并身不发生，本章的机制不启动。判断这一条的方法很简单：**问执行的作用面与判断的读取面是不是同一个面**。

五、**本章没有处理、也不打算处理的一个洞**：并身与病程自愈的交互。本章一直把二者当作可加的两部分，而临床常识告诉我们它们不是可加的——一个正确的干预往往是通过放大自然自愈来起效的，此时“哪一部分是药推的、哪一部分是自己走的”在原理上就不可分，不只是记录上不可分。本章的留底率对这个交互项无能为力：它只能说明“这次判断有多少依据来自未被自己动过的地方”，说不出被动过的那部分里两种来源各占多少。这是本章交出去的一个洞，而且是一个大洞。

六、**第二个洞**：本章假定“首方之前”是一条清晰的界线，而在真实临床上，病人在来看这一位医生之前往往已经吃过别人的药、自行服过成药、接受过其他系统的治疗。严格说，“未涉对照”在真实世界几乎不存在，本章所说的留底率其实是“未被本次判断涉及”的比率，不是“未被任何判断涉及”。这使留底率成为一个相对量而非绝对量。承认这一点，比假装它是绝对量诚实。

十七·证伪条件、样本纪律、伦理边界，与一条写死日期的赌注

十七之一·最强的竞争解释

在给出证伪条件之前，必须先把那句最难反驳的话摆在明处：

“这其实不就是病情自然演变吗？”

以及它的两个变体：“这其实不就是疗效吗？”和“这其实不就是交叉试验里的残留效应吗？”

一条证伪条件如果不能排除这三句，它就只是在验证“本章描述的现象存在”，那不算数。以下条件按能否排除竞争解释来设计。

十七之二·证伪条件

条件一（核心，机制证伪）：控制病程时长、基线严重程度、自然波动幅度与年龄后，若“上一版治法对当前主要证候要素的改写强度”（可用方剂中针对该证候要素的药味占比、剂量强度、疗程长度构造一个复合指标）与“复核结论相对上一版的偏移量”之间**不存在剂量-反应关系**，则本机制为伪——届时复核偏移应归因于病程本身而非执行。

条件二（排除残留效应）：在停药间隔长短不同的复诊病例中，若偏移量随停药间隔单调回落至零，则本判断被残留效应吸收。若偏移量收敛于一个不为零的平台，且平台高度与改写强度相关，则本判断成立。

条件三（排除知情机制，即循环效应）：在完全不知情的病例组（婴幼儿、语言不通者、意识障碍者）中，若并身量显著低于知情组，则本机制应归于循环效应。

条件四（低成本、可立刻实做）：所需数据在多数医院的既有记录里已经存在。调取任一家中医院近三年的门诊电子病历，按第九节的定义逐份计算留底率，若留底率与“复核结论偏移量”之间不存在负相关，则本章的核心读数无效。这一条不需要任何前瞻设计、不需要伦理审批之外的额外资源，一名研究生三个月可以完成。

条件五（正向读数，顺序预测）：在推行“首方前固定留底”规范的机构里，复核结论的医生间一致性上升，应当出现在规范推行之后的第1-两个复核周期，而不是与培训同步出现。若一致性的上升与培训同时发生、且在留底数据实际可用之前就已完成，则改善应归于培训效应而非本机制。**顺序比相关难被巧合满足。**

条件六（分层引用与分层证伪）：本章的主张分三层，可以分层引用、分层推翻。第一层是“并肩”这个构念本身（最强、最易倒）；第二层是那张两轴四格的辨别表（作为分析工具，即使第一层被推翻，第③格与第④格的区分仍然可用）；第三层是一条不依赖前两层的经验主张——“**临床复诊病历中，首方之前书写的记录占比极低，且该占比与病例管理规范程度负相关**”。第三层最稳、最便宜、也最容易检验；若连第三层都被推翻（即发现规范管理的病例反而留底率更高），本章全盘作废。

十七之三·样本纪律

上述条件涉及跨案例比较的，**样本不得少于六个同质单元，且在关键混淆变量上同质**。可用的同质单元包括：同一医院同一科室的多个医生组、同一方证的多个病例队列、同一区域内的多家同级医院。**严禁拿两家医院或两个地区的对比充数**——两边在病种构成、医生资历、信息化程度、病历书写规范上处处不同，所谓“控制掉某变量”在方法上站不住。这类比较只能作为启发式例证，不能充当证伪检验。

十七之四·伦理与证据边界

留底率是一个比率，一个比率一旦被提出，几乎必然会被拿去考核人。因此以下三条必须写死，缺一条本章不宜被引用：

一、**这个读数指的是位置，不是人**。留底率低说明“这一次复核站在软地上”，不说明这位医生水平差。急症、初诊即需用药、病人自行服药在先，都会使留底率低而临床决策完全正确。

二、**不得为了把这个数做好看而推迟必要的治疗**。任何以“提高留底率”为目的的停药、延迟给药、延长观察窗，都是不可接受的。第十二节第五条已经写明：可用的观察窗只能取自临床上本来就存在的自然间隔，不得人为制造。**这一条是硬约束，不接受任何“研究需要”的例外**。

三、**若已有机构据此读数对个人作出不利安排，必须公开编码规则并提供复核通道**。什么算一条“证据项”、什么算“结构上不受影响”，都有裁量空间，裁量规则不公开则不可被质疑。

十七之五·反噬预言

本命题一旦被制度采用，最省事的达标方式不是留底，而是在首诊时批量补录“基线四诊”——把复诊时才形成的判断倒填到首诊栏里，留底率账面上升而实质不变。更糟的一种是它会诱发延迟处方：为了留下一份未涉记录而推迟给药，把一个记账问题变成一次临床伤害。

我看不到出口。时点可以被伪造，而伪造的成本极低；唯一的抵抗是电子病历的写入时间戳不可篡改，而这是一个信息系统问题，不是一个临床问题，本章对它无能为力。把这一条写在这里，是因为一个看得到出口的反噬不是反噬，只是待办事项。

十七之六·一条写死日期的赌注

到二〇三三年十二月三十一日，至少有一部由国家级中医药行业组织颁布的证候诊断标准、证候疗效评价规范或中医病历书写规范中，出现如下强制条款：复核所依据的证据须标注其相对首次处方的时点，并单列首方前证据的比例。

什么不算命中，四条写死：

1. 仅出现在指南讨论稿、专家建议或学术论文中，未进入颁布文本的，不算；
2. 仅要求“重视基线记录”“完善初诊资料”而无时点标注字段与比例统计要求的，不算；
3. 仅在科研项目登记表、临床研究方案模板中要求，而未进入临床诊疗或病历书写标准的，不算；
4. 由本章作者或其所属机构参与起草的文本，不算。

十八·本章自己是不是它所描述的那样东西的标本

按本章的判据，本章自己落在哪一格？

先算本章的留底率。本章的承重命题是在与三门学科的顶撞过程中形成的，而本章用来支持它的材料——化学的能力验证共识值、教育学的练习效应、艺术学的介入档案——全部是在命题成形之后被检索、被选取、被排列的。也就是说，本章所依据的证据里，几乎没有一条是在“首次判断”之前固定下来的。本章的留底率接近零。

这不是一句谦辞，它有具体后果，至少三条：

第一，本章所举的每一个例子，都可能是被本章这道指令改写过的。例如第六节把《伤寒论》的坏病读成“一套完整的复核程序”——那十二个字确实可以这样读，但它是否本来就该这样读，还是被本章的框架推到了这个位置上，本章自己分不出来。要分出来，需要一份在本章成形之前就已固定的对《伤寒论》那一段的解读记录，而本章没有留。

第二，因此本章最不该被引用的方式，是被当作“已经证成”的东西引用。本章能提供的只有第十七节条件四那一条——一个便宜的、任何人可以在三个月内独立做出来的清点。在那个清点被独立完成之前，本章的地位是一个待检验的读数，不是一个结论。

第三，本章正落在自己所指认的第④格里。取不到未涉对照（命题与材料同时生成），而执行的产物并入了对象（那些例子已经被这个框架整理过一遍）。而且它有第④格的全部三条签名：短期指标良好（读起来自洽、例证丰富）、失效不产生信号（若它错了，读者不会在文中任何一处看到异常）、自我加固（框架越精细，反例越容易被读成框架的一个特例）。

对此本章只能做一件事，而且已经做了：在第十七节把最便宜的那条检验单独立出来，并写明它不需要本章作者参与。一个落在第④格里的判断，唯一的自救方式是把可以推翻自己的那件事做得足够便宜。

最后一句，关于这一整批文章。本系列近年的几篇——包括本章——共有一个形式：找出一样“发生了而没有产生条目的东西”。这个形式很有生产力，而生产力太高本身就是一个警号：一把在任何门上都拧得动的钥匙，通常不是钥匙，是撬棍。本章能给出的唯一防线是第十四节末尾那句话：若有人能证明这个形式在任意领域都能开出一样看似深刻的东西，那么这一批文章都要一起重估，本章排在被重估的第一位。

十九·伦理与证据边界声明

本章提出的留底率是一个结构读数，不是绩效指标。禁止用于个人考核与晋升评价；禁止以提高该读数为目的推迟或减少必要治疗；据该读数作出的任何不利安排，须公开编码规则并提供复核通道。本章所举临床情形均为类型化描述，不涉及任何具体病例。本章不构成任何诊疗建议。

注释

1. 本章所称「证」指中医辨证论治体系中的证候判断，不含西医意义上的诊断；所称“复核”指对已确立之证在后续诊次中的再次判断，不含同一诊次内多位医生的同期评判（后者为一致性问题，不在本章范围）。
2. 第二节对三门学科各自复核范式的描述取其主流形态，不代表该学科内部无争议；三家之间的冲突为本章所整理，各家并未如此互相表述。
3. 第六节对《伤寒论》“坏病”条文的读法为本章所作，仅取其程序结构，不涉及该条文的训诂与版本问题。
4. 第九节留底率的定义中，“证据项”的划分粒度会显著影响读数；本章建议以病历中可独立成立的记录条目为单位，并在任何比较研究中预先固定粒度规则。
5. 第十七节赌注的判定以颁布文本的正式发布日为准。

第五章 摊差

——一个证的复核为什么必然产出一份无人认领的分歧——论验证程序的终止性要求，与分歧留痕的位置

提要 一个证被确立之后要被复核，复核要给出结论，结论要写进病历。这条流程里有一个环节从未被记账：**复核过程中出现过的分歧，在结论产生的那一刻被处理掉了，而处理的形式是消灭而不是保存。**主任定了，学生的看法不入病历；共识达成了，少数意见不入标准；信度提上去了，被删掉的条目不入量表。本章把这份被消化掉的分歧命名为**摊差**——像会计上把一笔无法归属的差异摊进期间费用，摊完之后它不再作为一笔存在。本章的承重命题是：一个证的复核，不是把这个判断重新挂到一个更硬的锚上，不是让另一个人独立再判一次看是否一致，也不是检查从主张到证据的论证链上有没有断点，而是一次把新旧证据一起重新配权的结算；这次结算必然产出一份任何一方都消解不了的分歧，而这份分歧随即被吸收进“不确定度”“共识”“已接受”三类容器之一，从此不再作为一样东西存在。因此一个证算不算数，不取决于它被复核时是否通过，取决于复核所吃掉的那份分歧有没有留下技术内容。材料取自物理计量学的全局平差与扩展因子、心理测量的评分者一致性与共识程序、工程学的安全论证与开放项关闭三处成熟范式的互相顶撞；推翻它们共有前提的那件证据来自物理学自己——基本常数的推荐值不是某一次测量的结果，是全部历史测量在一次联立平差里被同时重新加权的结果，而输入数据互相矛盾时，矛盾被一个扩展因子摊进不确定度。本章给出一个可在四个领域并排清点的读数——**存异率**：一次复核中出现过的分歧条目里，在最终归档文件中仍可检索到其技术内容的占比；并给出一句不含情态词的问话：**这次复核归档的文件里，“未采纳意见”有没有一个字段？上一次填的是是什么？**结论是：复核不产出真相，它产出一份残渣；一个验证程序的可信度，写在它如何处置这份残渣上。

一·结论产生的那一刻，分歧去了哪里

先摆一个场景，它每天在发生，而且没有人认为它需要解释。

一位病人，两位医生看。主治判为肝郁脾虚，进修医生判为湿热中阻。两人各说各的依据：一个抓的是脉弦与情志诱因，一个抓的是舌苔黄腻与口苦。讨论十分钟，主治的意见写进病历，开方。

三天后有人调阅这份病历。他看到的是：**肝郁脾虚。**

那十分钟去了哪里？

它不在病历上。它不在处方上。它不在任何一份可检索的文件里。如果那位进修医生第二天调走，这份分歧就彻底不存在了——不是“被记在别处”，是**不存在**。而这份分歧里携带着真实的临床信息：有一位受过训练的观察者，从同一份四诊材料里读出了一个不同的落点，并且能说出依据。

同样的事在更高的层级上以更正式的方式发生。一部证候诊断标准的制定，通常走德尔菲法：几十位专家匿名打分，几轮反馈，把共识度未达阈值的条目删掉，剩下的写进标准。**被删掉的条目去了哪里？**它们通常只在制定说明里留下一句“经两轮咨询，共识度未达 0.8，予以删除”。删掉哪几条、每条被谁以什么理由反对、反对的临床依据是什么——不在标准里，往往也不在任何公开文件里。

而这里有一件值得停下来的事：**共识度低的条目，恰恰是最难判、最依赖经验、最需要辨别力的那些条目**。一条“面色萎黄”人人都能判，共识度自然高；一条“神情敛而不散”共识度必然低——而后者可能正是老医生真正在用的那一条。**标准化的标准做法是删掉后者**。

再看一层：证候疗效评价量表的编制。心理测量学教给我们的规范流程是，计算每个条目的区分度与信度贡献，删掉拉低内部一致性的条目。同样的机制，同样的结果：**留下来的是好判的，删掉的是难判的**。而信度报告上写的是：Cronbach's $\alpha = 0.91$ 。

这三层——一次会诊、一部标准、一张量表——是同一个动作在三个尺度上的重复：**复核产生了分歧，结论消灭了分歧，账面上呈现为一致**。

于是问题可以被写成一句话：

一个证如何被确立并被复核，才算数？

这个问句里有一个被默认的答案形状：算数=通过复核。而“通过”意味着复核给出了一个确定的结论。本章要说的是，**“给出确定结论”这个要求本身，是分歧被消灭的机制**；因而“通过了复核”这件事所携带的信息，比人们以为的少得多——它不告诉你判断有多可靠，只告诉你**这套程序有多擅长吃掉分歧**。

要看清这一点，在中医内部转不出来：中医没有“未采纳意见”这个字段，也没有关于这个字段的争论——它连争论都没有，因为这个位置上是空的，空得没有人注意到。所以本章从外面借三副眼睛。三门学科各有一套跑了几十年到一

百多年的成熟验证范式，三套范式互不兼容，而且冲突不在“谁更严格”，在它们各自把复核的终点锁在了不同的地方。

二·三种成熟的验证范式，与它们各自锁死的那个终点

二之一·第一家：物理学（计量学）——算数落在能不能接上那条链

一个测量结果算数，是因为它可以沿一条不断裂的比较链，一路追溯到一个被共同体定义的锚。

这就是溯源性。一台电子天平算数，因为它用一组砝码校准过；那组砝码算数，因为它送到省级计量院比对过；省级的标准算数，因为它接到国家基准；国家基准算数，因为它接到国际定义。二〇一九年国际单位制的重新定义，做的正是把七个基本单位的锚从器物移到基本常数上——从此千克不再由巴黎那块金属圆柱持有，而由普朗克常数的一个约定值持有。

这一家的终点是那条链所接到的东西，也就是整套定义与比对体系本身。落点不在读数上：一个读数再精确，若无法追溯，在计量意义上不算数；一个读数不那么精确但溯源完整、不确定度评定规范，是算数的。这一家的复核动作叫**复现性验证与周期检定**——把仪器重新接回链上。

它能成立，靠一个不写在正文里的前提：那个锚在体系之外，不参与被测量者的生成。普朗克常数不会因为你今天测了什么而改变。锚的外在性，是这一整套东西的地基。

二之二·第二家：心理学（心理测量与临床诊断）——算数落在当下的显现符不符合清单

一个诊断算数，是因为多位受过训练的独立评判者，面对同一位对象，给出**一致的判断**。

这一家在二十世纪七八十年代经历过一次决定性的转向：放弃对病因与机制的追问，把诊断标准写成可操作的清单——过去两周内，符合以下九条中的五条。这个转向的动机极其明确：在没有生物标志物、没有外部效标的领域，**唯一可及的“算数”就是一致性**。于是评分者间信度成为核心指标，kappa成为通用货币，结构化会谈成为标准工具。

这一家的终点是当下可观察的显现。不问为什么会这样，只问现在符不符合。复核动作是再来一次结构化评估，看两次的判断是否落在同一格。

它能成立，也靠一个前提：分歧是误差。整套信度理论建立在一个分解上——观察分数 = 真分数 + 误差；评判者之间的不一致被归入误差项，而误差是要被减小、被平均掉、被控制的东西。分歧在这一家的语法里，没有“携带信息”这个可能。

二之三·第三家：工程学（安全论证与验证确认）——算数落在论证结构上有没有断点

一个系统算安全，不是因为它测出了什么数，也不是因为它跑了多少小时没出事，而是因为存在一条从主张到子主张到证据的论证结构，且每一处断点都被处理。

这就是安全论证。它的形状是：我主张这个系统在这些条件下是可接受的；这条主张分解为若干子主张；每条子主张由证据支持；证据与子主张之间的推理步骤本身要被写出来并接受质疑。评审的动作是找断点：哪一步的推理没有支撑、哪条证据不足以支撑那条子主张。工程学同时区分两件事——验证是“我造对了吗”（符合规格），确认是“我造的是对的东西吗”（满足需求）。

这一家的终点是那条论证链本身。复核动作是再评审：把论证结构重新走一遍，看有没有新的断点，看开放项有没有关闭。

它能成立，靠的前提是：论证必须可关闭。一份安全论证若永远有开放项，系统就永远不能交付。整套制度的存在理由，是让判断能够终止在一个“可以交付”的状态上。

二之四·三家在打架，而且打得没法调解

物理学与心理学：物理学的整套体系是为了摆脱人的判断而建的——二〇一九年改革的全部意义就是把锚从“某一件东西、某一群人手上”移到“一个约定的定义”上。在这一家看来，共识不能充当锚：一百个人一致地测错，仍然是错。心理学的整套体系恰恰以共识为效标：在没有外部真值可及的领域，kappa就是它能拿到的最好的东西。两家对“共识能不能作为算数的依据”给出不相容的归属。若物理学对，心理学的整个诊断体系在计量意义上不算数；若心理学

对，物理学花一百年建溯源体系是自找麻烦。这不能靠判谁对化解——两家根本不在同一个语域里作答。

物理学与工程学：这一对是最硬的一对。物理学每天在做的事是**把每一处不确定都量化，一路合成，最后报告一个带不确定度的值**——它要求整条链上不许有任何一环靠判断。工程学恰恰把这件事诊断为病理：安全论证的经典批评之一，就是**用一个概率数字掩盖论证的断点**——一个算出来的失效率如果建立在无法验证的独立性假设上，它比一句诚实的“这一处我们没有把握”更危险。反过来，工程学每天在做的事——用论证结构、用专家判断填补证据缺口、用“合理可行”来关闭开放项——在计量学是完全不可接受的：**一条溯源链上只要有一环靠判断，整条链就断了**。一门学问的日课，正是另一门学问诊断为病根的那件事。

心理学与工程学：心理学要求判断落在当下可观察的显现上，刻意剔除机制——这是它的去理论化立场。工程学明确拒斥只有显现没有机制的证据：一份安全论证如果只写“我们测试了十万小时没出事”而不写“为什么它不会出事”，会被评审直接驳回，因为那不构成论证，只构成观察。前者说：不要问机制，看清单；后者说：只有清单没有机制，不算数。不能同真。

三家各锁一个终点：**那条溯源链、当下的显现、那条论证结构**。三个终点互不隶属，谁也折算不成谁。而正是在它们争夺终点的时候，三家脚下踩着同一块地。

三·三家共同假定的那件事：结算是可以完成的

三家争的是一个判断的效力该由哪一个终点来结算。争得很凶，而争的形式本身暴露了一条谁也没说出口的假定：

结算可以完成。也就是说：走到那个终点，判断就被结清了；剩下的只是维护，不再有未决。

把这句话念给三家听，三家都会说“这还用说吗”，而且三家都会立刻指出自己有装置保证它——

·物理学说：所以我们有不确定度评定。凡是没能被消除的，都被合成进那一个数里，报告出来就完了。

- 心理学说：所以我们有共识程序。评判者不一致的地方，走裁决或取多数，得到一个最终值就完了。
- 工程学说：所以我们有开放项关闭制度。评审提的每一条质疑，逐条关闭，清零就完了。

三套装置，做的是同一件事，而且是同一个动作：**把未决转化成已决**。不确定度是一个筐，共识是一个筐，“已接受”是一个筐。三只筐的形状不同，功能相同：**它们让程序能够终止**。

本章把这条假定叫做**终止性要求**——一个验证程序必须能够结束，而结束意味着所有未决必须被处理掉；于是在制度层面上，“处理”与“消灭”是同一个动作。

现在把这一条搬到辨证论治上，它当场断裂，而且断的方式与另一篇同题文章所指出的完全不同。

- **没有外在的锚**。证没有金标准。没有一个不参与临床判断生成过程的东西，可以让辨证结论去接。有人说金标准是疗效——但疗效的评价同样是辨证的一部分（证候积分由谁判？还是那位医生），而且疗效滞后于判断，无法用于当次结算。
- **共识不构成效标，但它是唯一可及的东西**。于是中医的现代化路径几乎必然地走向了心理测量的模子：算 kappa，办德尔菲，编量表，提信度。这条路走得很认真，而它的每一步都在强化那只筐。
- **论证结构无法关闭**。辨证的推理链上，从四诊到八纲到脏腑到证型的每一步，都包含大量无法被证据单独支撑的判断步骤；若按工程学的标准逐条要求支撑，整条链会停在第一步。

三家的装置在这里全部失效，而失效的方式不是它们不工作，恰恰是它们**工作得再好**：中医现代化引进的正是这三只筐，而在一个没有锚、没有效标、论证链无法关闭的领域里，筐是唯一能让程序终止的东西。**筐用得越好，被装进去的东西越多**。

于是本章要处理的问题变得具体了：**在没有外在锚的地方，复核这个动作实际上在做什么，以及它每次做完之后剩下的那一份，去了哪里**。这是一个关于路径与去向的问题，不是一个关于概念的问题。答案不会是“证的本质是什么”，答案会是一条次序，和次序尽头的一个位置。

四·物理学自己的一件材料：没有任何一个常数是单独确立的

要推翻一条三家共同踩着的假定，从外面搬理由来没有用——那只是加入争论，成为第四家，三家可以各自退回阵地。要让三家退不回去，推翻它的那件材料必须是三家之一自己已经掌握、已经公开、只是从来没往这个方向用的事实。

这件材料在物理学手里，而且它每四年更新一次，全世界都在用。

基本物理常数有一套国际推荐值。人们通常以为，普朗克常数的推荐值是某一次最精确的测量的结果。**不是**。它是一次**全局最小二乘平差**的输出：把全世界所有相关的测量结果——瓦特天平、硅球计数、约瑟夫森结、量子霍尔电阻、光谱学数据——连同它们各自自报的不确定度，一起放进一个联立方程组，同时求解全部常数。输出的不是一个数，是一套**互相关联、内部一致的值**，每个值带一个不确定度，值与值之间带相关系数。

这意味着三件事，每一件都在动摇上面那条共有前提：

第一，没有任何一个常数是单独确立的。它们是被一起解出来的。一个常数的值里，含着所有其他常数的测量信息。

第二，每一次新测量进来，所有旧值都会被重新算一遍。不是“新的推翻旧的”，是**全部输入被同时重新加权**。某一次十年前的实验，可能因为今年的一次新测量而被下调权重——而它自己什么也没变。**复核在这里根本不是对某个判断作裁决，是一次全局配权。**

第三，也是最要紧的一件——当输入数据互相矛盾时，矛盾被一个扩展因子摊掉。平差之后如果卡方检验说输入之间不自洽（也就是有几组实验的结果在各自宣称的不确定度下无法同时为真），标准做法是给全部输入的不确定度**乘上一个大于一的因子**，一直乘到卡方可以接受为止。这个因子在计量学里有一段很长的历史，早在二十世纪三十年代就被讨论过。

请看清楚这个动作在做什么：

有几组实验互相矛盾。矛盾没有被解决——没有人查出是哪一组错了、错在哪里。它被转化成“所有人的不确定度都比自己以为的大”，摊进了输出的那个不确定度里。摊完之后，输出是一套自治的推荐值，而那份矛盾不再作为一份矛盾存在。

物理学人当然知道这一条，而且把它写在方法说明里，还配了一整套告诫——扩展因子过大说明存在未识别的系统误差、应当优先查找而不是简单放大、必要时应剔除可疑输入。但这整套讨论回答的是“如何给出一套自治的推荐值”。它从来不被拿来回答“复核的结构是什么”。这正是它能推翻共有前提的原因：在它自己的领域里，它回答的是另一个问题。

把它读成一条关于验证结构本身的证据，得到的是：

在没有外在锚可及的场合，复核不是对某一个判断的裁决，是一次把新旧证据一起重新配权的结算；而这次结算必然剩下一份任何一方都消解不了的分歧，这份分歧被一个吸收装置转成“已处理”。

这一句一旦成立，共有前提垮掉：结算不能被完成，只能被终止；而终止的方式是把未决转化成一种不再是未决的形态。

还有第二件材料，来自物理学的另一处，量级更大：哈勃张力。两条原理完全独立的路径——近距离的距离阶梯与早期宇宙微波背景加上标准模型——各自的精度这些年不断提高，而两者给出的膨胀率不但没有收敛，分歧随精度提高而放大。这件事在物理学自己那里被读作“可能存在新物理”，是一个物理问题。把它读成一条关于验证结构的证据，它说的是：“原理独立的多条路径必然殊途同归”不是一条被验证的结论，是一条被假定的前提；而这条前提在精度足够高时会失效。

值得注意的是，在这个张力持续存在的年份里，宇宙学的下游工作照常进行——论文按需要选一个值用，标注取自哪一边。一个未结清的判断，照样在支撑全部下游工作。这就是第③格（本章第七节）的一个现成标本：分歧巨大，公开未决，而世界照转。物理学做到了这一点，而它是极少数做到的领域之一。

第三件材料在心理学手里，可以顺带一提，因为它同样有效：德尔菲法的第二轮反馈会把上一轮的意见分布呈现给参与者——四分位间距、离群者被要求说明理由。这一步是它整个方法的核心。但呈现的是分布，不是内容：参与者看到“你的打分在离群区间，请说明理由”，而最终标准里既不留分布，也不留理由。心理学自己知道分歧携带信息（否则不必反馈），但它使用分歧的方式是让分歧自己消失——反馈的目的是促成收敛。

到这里可以把话说到底：三家看不见这一层，不是因为它们疏忽，是因为它们各自的核心问题恰好使这一层成为不必要的。物理学的核心问题是“这个数是

多少”，它有扩展因子这个装置，装置有效，因此不必往下看那份被摊掉的东西是什么。心理学的核心问题是“这个判断可不可靠”，在它的语法里分歧的定义就是误差，误差不是要保留的东西。工程学的核心问题是“这个系统能不能交付”，交付要求关闭，**关闭就是它的目的**，而不是它的副作用。三家的盲区，是三家各自问题结构的影子。

五·摊差：一份被转化掉的分歧

现在可以给那样东西命名了。

一次复核，无论在哪个尺度上——一次会诊、一轮德尔菲、一次平差、一次安全评审——都会经历这样一个过程：多个来源的判断被放在一起，其中一部分互相支持，一部分互相矛盾；程序必须在某个时刻终止并给出结论；终止时，那些互相矛盾的部分被一个装置处理掉。

本章把被处理掉的那一份叫做**摊差**：

摊差——一次复核中真实出现过、任何一方都无法消解、在结论形成时被吸收到“不确定度”“共识”“已接受”三类容器之一、从此不再作为一份分歧存在的那一份未决量。

用会计的话说，它被**摊销**了：一笔无法归属到任何具体科目的差异，被摊进期间费用，摊完之后报表平了，那笔差异不再作为一笔存在——它不是被隐藏，是被**转化**。

要看清它是一样独立的东西，最快的办法是把它与四个相邻的东西分开：

它不是误差。误差有一个隐含的参照：真值。说某个判断有误差，前提是存在一个正确答案。摊差发生在**没有可及真值**的场合：两位医生的分歧不是“一个对一个错”，而是两个受过训练的观察者从同一份材料里读出了不同的落点，而没有任何东西可以裁决他们。**把摊差记为误差，是本章要指认的那次转化本身。**

它不是噪声。噪声的特征是随机、无结构、可以通过平均减小。摊差有结构：同一位医生反复在同一类条目上与同行分歧，这不是随机的；一部标准里被删掉的条目在临床辨别力上系统性地高于保留条目，这更不是随机的。**平均不会让摊差变小，只会让它更快地消失。**

它不是待办事项。 待办有一个负责人、一个期限、一条回来处理的路径。摊差没有——它被关闭的那一刻，就没有任何流程会再回来看它。工程学的开放项清单是本章所知的唯一近似物，而它的实际命运见第十五节。

它不是默会知识。 这一条最要紧，因为它是最容易被拿来吸收本章的说法。默会知识指的是**从来没能被说出来的东西**——老医生手上那点说不清的东西。摊差恰恰相反：**它被说出来过**。那位进修医生说了他的依据，那位德尔菲专家写了他的理由，那位评审员提了他的质疑。**它不是说不出来，是说出来了然后被吃掉**。这一处分界后面还要用来构造一条极硬的证伪条件。

摊差的三条性质，本章后面全部要用：

1. **它是被正确运行的程序所产生的。** 不需要任何人违规，不需要任何人压制异议，不需要任何社会性的施压。程序按设计运行，结论必须唯一，唯一就意味着吸收。**摊差是合规的产物，不是违规的产物。**
2. **它的量与程序的质量指标正相关。** 一致性越高、共识度阈值越严、开放项清零越彻底、不确定度报告得越干净——被吸收的越多。这一条反直觉，是本章的枢纽，第八节专门处理。
3. **它不可归咎，因而无人有动机记录它。** 没有人做错事，因此没有人需要为它免责或追责。记录制度是被责任驱动的，于是没有人为它开一栏。

三条合起来是一个少见的形状：**一样确实出现过、量可能很大、对下一次判断有真实价值、而且没有任何一方做错事的东西，因此它在所有账本上都不设字段。**

六·三只筐：不确定度、共识、已接受

摊差不会凭空消失，它必须被装进某个东西里。三门学科各提供了一只筐，形状不同，功能一致。看清这三只筐，是本章最实用的一节。

第一只筐：不确定度。

物理学的做法是把无法消解的分歧转化成一个数——“我们的不确定度比原来以为的大”。这只筐的高明之处在于它诚实：它没有假装分歧不存在，它把分歧的**规模**保留下来了。它的代价在于它丢掉了分歧的**内容**：一个被扩展因子放大三倍的不确定度，告诉你“有一份没解决的矛盾，大约这么大”，不告诉你“是哪几组实验、在哪个方向上、因为什么”。**规模留下了，方向和理由摊掉了。**

在中医里对应的是什么？对应的是“**证候诊断存在一定主观性**”这句话。这句话出现在几乎每一篇证候研究的局限性讨论里，它就是那个扩展因子——它承认有一份未消解的东西，给它一个笼统的规模（“一定”），然后继续。

第二只筐：共识。

心理学的做法是让分歧通过程序收敛：多轮反馈、离群者说明理由、达到阈值、取共识值。这只筐的高明之处在于它**做工作**——它不只是掩盖，它确实把一堆不可用的判断加工成了一个可用的结论，而且有证据表明经过共识程序的诊断预测效度更高。它的代价在于：**加工的残渣不留账**。那些被要求说明理由的离群意见、那些没达到阈值的条目、那些在第二轮改了口的人改口的理由，全部不进入产物。

在中医里对应的是什么？对应的是**德尔菲法制定的每一部证候标准**，也对应**每一次由资历决定结论的会诊**。后者是共识筐最省事的形态：不必投票，主任说了算。

第三只筐：已接受。

工程学的做法是把无法解决的东西转化成**一个被授权的接受动作**——残余风险已评估、合理可行范围内已降至最低、开放项予以关闭。这只筐的高明之处在于它有**责任人**：接受是一次签字，签字有名字有日期。它的代价在于：**签字记录的是接受这个动作，不是被接受的那件事的技术内容**。十年后翻出那份文件，你看到的是“某年某月某人接受”，看不到“当时的争议是什么”。

在中医里对应的是什么？对应的是“**结合临床，随证治之**”这类收尾语。它在临床上是有用的、必要的、正确的——它把一个无法在此刻确定的判断合法地移交给下一次。而它同样只记录移交动作，不记录被移交的内容。

三只筐并排看，性质很清楚：

	保留了什么	摊掉了什么	有无责任人
不确定度	分歧的规模	分歧的方向与理由	无
共识	一个可用的结论	少数意见的内容	通常无
已接受	一次授权动作与责任人	被接受之事的 技术内容	有

三只筐各留下一样东西，各摊掉一样东西，而没有任何一只筐保留分歧的技术内容。这不是巧合：保留技术内容意味着程序不能终止，而三只筐存在的全部理由就是让程序能够终止。

这也就是为什么本章要打的不是某一只筐，是那条“终止即消灭”的等式。

七·两根轴与四格

收成一张辨别表，需要两根轴。

第一根轴：这次复核产出的分歧量。 复核过程中出现过多少条互相矛盾、任何一方都消解不了的判断。

第二根轴：这些分歧的去向。 是留下了技术内容，还是被吸收进三只筐之一。

	留下技术内容	被吸收进筐
分歧量低	① 真收敛 ——本来分歧就少，有的那点也记着。罕见而理想。	② 表面收敛 ——分歧本来就少，吸收也没吸掉多少。多数常规检验落在这里，无害。
分歧量高	③ 公开未决 ——分歧巨大且全世界看得见，下游照常工作，各自标注取值来源。哈勃张力在这一格；英美法判决书的异议意见制度也在这一格。	④ 靶格

第④格：分歧量高，而且全部被吸收，账面上呈现为高度一致。

八·靶格：所有质量指标都在改善的那一格

第④格不是一个坏格子。它是一个看上去最好的格子，而且是被整个质量改进事业主动追求的那个格子。这是它值得单独写一篇文章的全部理由。

一个落在第④格的复核，形式审查全部通过：

- 评分者间一致性达标，kappa 高；
- 共识度达到阈值，标准得以颁布；
- 开放项清零，评审通过；
- 报告的不确定度小，结论明确；

- 病历上结论唯一，无涂改，书写规范；
- 若有质控，“诊断符合率”漂亮。

没有任何环节报错。而恰恰在这样一次复核里，被吃掉的分歧可能是最多的。

三条签名，逐条对照：

签名一：它在短期指标上表现为改善，而且是因果性的改善。这一条比通常的“看起来好”更强。提高一致性的标准做法有三样：培训评判者到趋同、删掉难判的条目、设立裁决程序。这三样每一样都是吸收装置。也就是说，指标的改善与摊差的增加不是并存关系，是同一个动作的两面。一个 kappa 从 0.45 提到 0.92 的量表，它提高的那 0.47，主要不是来自“大家看得更准了”，是来自“难判的地方不再需要判了”。

签名二：它的失效不产生任何信号。如果第④格的复核给出了一个错误的方向，谁会知道？没有人。因为被吸收的分歧不再是一样东西，它不会报错，不会出现在任何异常清单上。下游拿到的是一个干净的共识值，干净得看不出它是怎么来的。唯一会暴露它的场合，是很久以后有人从别的方向撞上同一个问题，而那时通常已经无法回溯当初的争议。

签名三：它自我加固，而且越正规化越严重。这一条最要命。一切质量改进的方向都是提高一致性、减少歧义、明确结论、清零开放项。每一项改进都在加强吸收装置。一个尚未标准化的领域，分歧明摆着，谁都看得见；标准化的第一步就是把这些分歧处理掉——而处理的形式是消灭。因此第④格是标准化的自然终点，不是标准化的失败。

这里必须把话说清楚，否则本章会被读成反对标准化：**吸收是必要的。**一个不能终止的程序无法交付，一个没有结论的会诊无法开方，一部没有共识的标准无法颁布。本章不主张停止吸收，本章主张的只有一件：

吸收的时候，把被吸收的那份东西的技术内容单独留一笔。

这是一个记账要求，不是一个决策要求。它不改变任何结论，不推迟任何交付，不动摇任何权威。它只要求：**结论旁边留一栏，写清楚这次吃掉了什么。**

九·存异率：一个四领域量纲不同而比率相同的读数

一个判断若停在“有这么一样东西”，它还不能被检验。要能被检验，需要一个可以事后清点的数。

本章给出的读数是**存异率**：

存异率 = 一次复核中出现过的分歧条目里，在最终归档文件中仍可检索到其技术内容的条目数 ÷ 该次复核中出现过的分歧条目总数。

“技术内容”的定义要写死，否则这个读数会被一个空字段架空：**可检索到技术内容**，指的是能读出“分歧在哪一条、持不同意见者的判断是什么、依据是什么”三样。只写“存在不同意见”或“经讨论未采纳”的，不计入分子。

分母的来源必须交代，因为分母恰恰是被吸收掉的那个东西。它可以从**过程记录**取得，而过程记录在四个领域里都已经存在：

- **物理计量学**：平差的输入数据集是公开的，卡方贡献、被下调权重的条目、扩展因子的取值都在方法说明里。**存异率 = 那些造成不自洽的输入条目中，在最终推荐值文档里被逐条指名并说明其分歧方向的比例。**这一家的存异率在四家中最高——它至少公布了输入清单与因子。
- **心理测量**：多评分者的原始评分表存在，德尔菲各轮的原始打分与理由栏存在。**存异率 = 共识前存在分歧的条目中，在最终量表或标准文本（含编制说明）里可读出分歧内容的比例。**这一家的存异率通常极低。
- **工程安全论证**：评审记录、开放项清单、关闭记录存在。**存异率 = 评审中提出的质疑中，关闭记录里写有技术性关闭理由（而非“已澄清”“可接受”）的比例。**
- **中医临床与标准制定**：会讨论记录、师承带教的辨证分歧、同一病例多医生辨证的原始记录、标准制定的咨询轮次记录。**存异率 = 上述分歧条目中，在最终病历或标准文本里可检索到技术内容的比例。**

四处的单位分别是“输入条目”“量表条目”“开放项”“病历/标准条目”，量纲毫不相干；比率是同一个东西：**一次复核吃掉了多少分歧而没有吐出痕迹。**

这个读数要立住，必须过四关：

第一关，无量纲。是比率，四个领域可以并排放在同一张图上比较。

第二关，可事后清点。不需要新仪器、不需要前瞻设计，只需要两样已经存在的东西：**过程记录与归档文件。**拿一份会诊记录与对应的病历，逐条比对；拿

一部标准的咨询轮次原始表与颁布文本，逐条比对。多数机构两样都有，只是从未被并排读过。

第三关，把一个印象变成一个数。“会诊的分歧其实都没留下”是一句人人有感觉、无人能拿它做事的印象。0.03 是一个数，可以排序、可以比较、可以进行设计。

第四关，与既有主要指标正交。这一关最容易漏，也最容易毁掉一个新指标——若它与某个现成指标高度相关，它会被立刻读成那个指标的代理。

存异率过这一关，而且是**反向过**的：

一部证候诊断标准，三轮德尔菲，共识度阈值 0.8，最终条目内部一致性 $\alpha = 0.93$ ，专家积极系数 100%，权威系数 0.89——**它的存异率可以是 0**。因为凡是没达到共识度的条目都被删除，而删除记录只写“共识度未达标”，不写谁反对、为什么反对、反对者的临床依据是什么。

>

另一份材料：某医院一次疑难病例讨论的完整录音整理稿，五位医生四种辨证意见，最终未能统一，病历上写“暂按肝郁论治，一周后视舌脉再议，另附讨论记录”——**它的存异率接近 1**，而它的“诊断一致率”是 0.2。

一致性指标越高，存异率越低；一致性指标越难看，存异率反而越高。这一条足以说明存异率不是任何现有质量指标的影子，也预告了一个必须正面处理的风险：它一旦进入考核，最省事的达标法是制造无意义的分歧。见第十七节。

必须写死一条：**存异率不是越高越好**。它是一个**结构读数**，不是绩效指标。一次真正没有分歧的复核，存异率的分母为零，这个读数不适用，而那完全正常。存异率说明的是“这次复核吃掉的东西有没有留下痕迹”，不说明“这位医生水平如何”，更不说明“分歧越多越好”。

十·一句不含情态词的问话

一个判断如果只能靠“应当”“充分”“真正”“恰当”来表述它的判据，它还没有落到可查的层面。本章的判据是一句可以直接拿去问归档文件与流程规范的问话，一个情态词也没有：

这次复核归档的文件里，“未采纳意见”有没有一个字段？上一次填的是什么？

它问的是事实，不是判断。回答它不需要任何人的意见，只需要翻文件。

把它在五个场合各跑一遍，答案互不相同——这一点很重要：如果五个场合答出同一个东西，那说明它只是承重命题的复述，不是一个能分辨情况的判据。

场合一，医院疑难病例讨论。答案通常是：有一份讨论记录，但它与病历分离存放，病历上的诊断栏没有指向它的链接。**分歧的技术内容存在，而检索路径不存在。**这提示了一种此前没想到的中间状态：**留了，但接不上**——它的实际效果接近于零，因为下一位调阅病历的人不会知道有那份记录。这一条是查出来的，不是从命题推出来的。

场合二，基本常数推荐值的更新。答案是：有。输入数据清单、卡方、扩展因子取值、被特殊处理的输入，都在方法说明里逐条列出。**存异率在四家中最高。**但同时暴露出另一件事：**列出来了，也很少有人去读**——下游使用者拿的是那个数，不是那份说明。**留痕的价值取决于有没有回读的场合**，这一条同样是查出来的。

场合三，一部行业标准的德尔菲制定。答案通常是：没有这个字段。编制说明里有“经三轮咨询，删除条目 X 条”，而 X 条是哪几条、为什么删，多数不列。**存异率接近零，且分母也无法从公开材料重建。**

场合四，一份安全论证的评审关闭。答案是：有开放项清单，每条有关闭记录，而关闭记录里技术内容的丰俭差别极大——从“经分析，风险可接受”到整整两页的分析。**同一个字段，填法决定一切。**这提示：设字段是必要条件，不是充分条件。

场合五，一次法庭合议。答案取决于法域：有些法域的判决书公开异议意见并署名，异议意见全文保留、可被引用、可能在几十年后成为多数意见；有些法域的合议庭评议笔录不公开，判决书只呈现统一意见。**同一个功能位置，两种截然不同的制度安排，都运行了很久，都能交付。**这是本章最有用的一处发现：**保留分歧的技术内容并不妨碍程序终止——已经有一整套成熟制度证明了这一点。**

五个答案互不相同，其中至少三个（留而接不上、留而无人读、同字段两种填法）不能从承重命题直接推出。

十一·复核路径的重排：那一笔该记在哪一步

现在可以正面回答那个问题了：一个证如何被确立并被复核，才算数？

现行的默认路径是两步：

确立 → 复核（通过／不通过）

本章主张它实际上是四步，而第三步现在没有位置：

确立 → 重新配权 → 摊差结算 → 出结论

四步各自的动作：

1. **确立**：四诊合参，判出证，形成一道执行指令。
2. **重新配权**：复核不是把上一版拿来审，是把上一版所依据的全部信息与新到的信息放在一起重新称量。这一步在临床上一直在发生，只是它被误读成了“审查上一版对不对”。误读的后果很实际：审查的语法要求给出“对/错”，而配权的语法给出的是“哪些证据的权重变了”。
3. **摊差结算**：在给出结论之前，把这次无法消解的分歧单独清点一遍——哪几条、谁的判断是什么、依据是什么。这一步现在完全没有位置。
4. **出结论**：给出唯一结论，据以处置。这一步必须保留，本章不动它。

三个要点：

第一，可插手的那一步是第三步，而且只有第三步。现行的全部改进努力压在第四步的质量上——提高一致性、统一标准、加强培训、明确结论。这些改进各自都对，而它们全部通过加强吸收来实现，因此每一次改进都在把第三步的空位压得更紧。这是本章与主流质量改进方向的正面分歧，必须说清楚：本章不反对第四步做得更好，本章反对的是在没有第三步的情况下把第四步做得更好。

第二，第三步不增加任何决策成本。它不要求推迟结论、不要求保留争议、不要求任何人让步。主任照样定，标准照样颁，方照样开。它只要求在结论旁边多一栏，写清楚吃掉了什么。一个记账动作，不是一个决策动作——这一点必须反复讲，因为它是这条主张唯一可能落地的理由。

第三，第三步的成本几乎全部落在“回读”上，而不是“记录”上。记一笔很便宜；让下一个人读到那一笔很贵。第十节场合一与场合二说明了两种失败方式：留了但接不上（检索路径断裂），留了但无人读（没有回读的场合）。因此

第三步的完整形态是：留一笔，并且规定在什么情况下必须回读它。最自然的回读触发条件有两个——**结论被推翻时、同一位病人第三次复核时**。

三门源学科的路径在这个框架下可以重新排列，各自有效性的来源也清楚了：

- **物理学**走的是“从被测量出发，经比对链，落到定义锚”；它把第三步做成了扩展因子与输入清单——它记了，只是记的是规模不是内容，而且下游不回读。
- **心理学**走的是“从当下显现出发，经共识程序，落到一致的判断”；它把第三步整个压掉了——分歧在它的语法里是误差，误差没有记录的资格。
- **工程学**走的是“从主张出发，经论证结构，落到可交付”；它把第三步做成了开放项清单与关闭记录——它设了字段，字段的填法决定它是第③格还是第④格。三家里，工程学离本章的主张最近，而它离得近的原因是它有责任人：接受要签字，签字的人希望日后能说清自己当时接受的是什么。

十二·七个可以立刻做的记账动作

一条判断如果给不出低成本的落地动作，它就只是一个说法。以下七条按成本从低到高排列，前四条不需要任何新增流程。

一、在会诊与疑难病例讨论的结论栏旁边，加一栏“未采纳意见”，格式固定三段：哪一条、判断是什么、依据是什么。这一栏允许为空，但为空必须是一次主动选择——即讨论中确实无分歧，而不是有分歧未记。这是本章全部主张里最便宜的一条。

二、把讨论记录与病历诊断栏之间建一条检索链接。针对第十节场合一那种“留了但接不上”的失败：分歧的技术内容已经存在于讨论记录里，缺的只是一个指针。这是信息系统的一行配置，不是临床工作量。

三、标准与量表的编制说明里，逐条列出被删除的条目及其被删除的原因，区分“共识度未达标”与“临床不重要”。这两件事被混为一谈是当前最大的一处信息损失：一条因为**难判**而共识度低的条目，与一条因为**不重要**而被大家一致认为可删的条目，命运相同而含义相反。分开列，成本为零。

四、把“结合临床，随证治之”这类移交语，改为带内容的移交：移交什么、移交给谁的判断、下次回来看什么。移交本身是正确的临床动作，本章不动它；本章只要求移交单上写内容。

五、给存异记录规定回读触发条件。至少两条：结论被后续推翻时必须回读；同一位病人第三次复核时必须回读前两次的存异栏。没有回读规定的留痕，实际效果接近于零。

六、借法庭异议意见的形式。有些法域的判决书公开异议意见并署名，全文保留、可被引用、可能在几十年后成为多数意见。这一形式对中医会诊完全适用，而且解决了本章最担心的一个问题——署名使留痕对留痕者有价值，从而不必依赖行政强制。它也解决了权威受损的顾虑：异议意见的存在从未削弱判决的效力，因为判决主文与异议意见在效力上是分开的。

七、对标准制定，保留咨询轮次的原始数据并在若干年后开放。这一条成本最高、阻力最大，但它是唯一能让“被删条目的临床辨别力”成为可研究对象的办法，也是本章第十七节那条最便宜的证伪条件所需要的数据来源。

七条里第一条与第五条最要紧：一条让摊差可以被清点，一条让它可以被读到。只做第一条不做第五条，得到的是一个没有人打开的抽屉。

十三·与最近的几种说法的分界

一个新名字最常见的下场，是被读者正确地认出来是某个旧概念的改名。要避免这个下场，只有一个办法：把最容易与它混淆的说法逐个请进来，各自说清它到哪一步，再说清同一份材料上两边会给出什么不同的判断。

先把本章的命题剥掉中医与计量的一切名词，只留形式：

一个验证程序在产出结论的同时消灭了它所处理的分歧，而消灭的彻底程度与该程序的质量指标正相关。

这个纯形式的结构，属于“越成功越失败”这一类反转结构，而这一类结构在管理学与组织社会学里早有名字——学习近视与成功陷阱（Levinthal & March 1993）、核心刚性（Leonard-Barton 1992）、目标置换（Merton 1940）。必须先点名最贴的那一个再往下走。最贴的是成功陷阱：组织因为擅长某件事而不断强化它，直至强化本身成为失效来源。分离线：成功陷阱说的是开发挤压探索，两边是同一类活动的两种配置，代价在未来显现；本章说的是结论

挤压分歧，两边不是同一类活动，代价在**可回溯性**上显现且立刻发生。判定形式：若加强一致性的机构在未来的创新能力上受损，则成功陷阱对；若其创新能力不变而旧结论一旦被推翻就无法查出当初争议，则本判断对。二者可以同时为真，互不蕴含。

以下逐位。

（一）扩展因子与 Birge 比（物理计量学，Birge 1932）

它到哪一步：它是本章的直接材料来源，必须先声明本章对它的关系是借用而非发明。它承认输入之间可能不自洽，并给出一个把不自洽转化为不确定度的操作。

分离线：它是一个**装置**，本章是关于这个装置的**账目位置**的判断。它回答“不自洽时怎么给推荐值”，本章回答“被它吸收掉的那份东西是不是一样东西、该记在哪里”。**本章不主张扩展因子是错的；本章主张凡有扩展因子处，必有一份未被记账的内容。**

判定形式：若在扩展因子的使用中，被放大的那些输入的分歧方向与理由都被逐条记录并可回读，则本章所指认的空缺在物理学不存在；若只记规模不记内容，则存在。**这一条可以直接去查文档，不需要任何新研究。**

（二）评分者间一致性与共识程序（心理测量，Cohen 1960 及后续）

它到哪一步：它把评判者之间的不一致定义为误差，并提供了度量与减小的整套方法。

分离线：它把分歧读成误差，本章说分歧携带信息。这不是“侧重不同”，是同一份材料上的相反判断——**一条 kappa 拉低的条目，信度理论判定为应删除，本章判定为应重点保留并记录理由。**

判定形式：取一部标准德尔菲过程中被删除的条目，与保留条目一起，在独立的病例集上考察其对临床结局或治疗选择的辨别力。若被删条目的辨别力不高于保留条目，则本判断错，信度理论的处置正确；若系统性更高，则本判断对。**这是本章最便宜、也最要命的一条检验。**

（三）德尔菲法（Dalkey & Helmer 1963）

它到哪一步：这是最容易吞掉本章的一位，因为它**明确处理分歧**——第二轮会把意见分布反馈给参与者，离群者被要求说明理由。它看上去已经做了本章要求的事。

分离线：它保留的是分歧的分布（四分位间距、离群标记），不是分歧的内容；而且它保留分布的目的是促成收敛，不是留档。轮次结束后，分布与理由通常都不进入产物。同一个动作，用途相反：德尔菲用分歧来消灭分歧，本章要求把分歧留给下一次。

判定形式：查任一部已颁布标准的德尔菲原始材料是否随标准公开；若公开且含理由栏，本章所指的空缺在此不存在。若只公开轮次数与共识度阈值，则存在。

（四）开放项与合理可行降低原则（工程安全）

它到哪一步：这是本章最近的方法学邻居，也是三家中唯一为分歧设了字段的一位——开放项清单、关闭记录、有责任人的接受签字。

分离线：它设了字段而不约束填法；它要求关闭而不要求关闭理由的技术内容；它更没有回读规定。本章相对它的增量只有两条：填法的最低内容要求，与回读的触发条件。这两条听起来微小，而第十节场合四说明它们决定了这个字段是有用还是空转。本章对这一位的关系不是取代，是收紧。明写这一点，因为把一位近邻写成稻草人是最容易被拆穿的做法。

判定形式：在同一批安全论证中，比较关闭记录含技术内容与不含技术内容两组，考察结论被后续修订时的可回溯性。若两组无差异，本章的收紧是多余的。

（五）集体思维（Janis 1972）

它到哪一步：它指出群体在追求和谐时会压制异议，导致决策质量下降。这是最容易被拿来吸收本章的社会心理学解释。

分离线：集体思维的机制是社会压力与从众动机，它要求有人想说而不敢说；本章的机制是程序性吸收，不需要任何人被压制——分歧被完整地说出来了，程序照样把它吃掉。

判定形式：在完全匿名、无社交互动、由算法自动合并的评判场合（如自动化多读者影像判读的仲裁流程、双盲德尔菲），若分歧的技术内容同样不可检索，则本判断对而集体思维不足以解释；若匿名化之后分歧留痕显著改善，则应归于集体思维。这一条可以直接做。

（六）发表偏倚与抽屉问题（Rosenthal 1979）

它到哪一步：不显著的结果不进入文献，导致文献整体偏倚。

分离线：抽屉问题是**未进入**——那些研究从来没被记录过；摊差是**进入了然后被消化**——它出现在过程记录里，然后在归档时消失。两者的干预点完全不同：前者要预注册，后者要留痕字段。

判定形式：若一个领域实行了完全的预注册（所有研究都进入记录），摊差是否消失？本章预测不会——预注册解决的是研究的存在，不解决评审分歧的去向。若预注册之后归档文件里的分歧留痕显著改善，则本章被抽屉问题覆盖。

（七）默会知识（Polanyi 1966）

它到哪一步：有些知识无法被完全表述，只能通过实践传递。这是解释“老医生的东西传不下去”最常用的说法。

分离线：默会知识指的是**从未被说出来的**；摊差指的是**说出来了、被记录过、然后在结算时被消化掉的**。二者的处置完全相反：默会知识需要师承、需要长期同处；摊差只需要一个字段。**把摊差误读为默会知识，会导致一个可以用记账解决的问题被当成一个只能用师承解决的问题——这个误读的代价极大，因为它把一个便宜的解法换成了一个昂贵的解法。**

判定形式：只统计曾以文字形式出现在过程记录中的分歧条目。若这部分同样大比例不可检索，则不是“说不出来”，是“说了被吃掉”。**这一条把默会知识彻底排除在外，因为它只用已经被说出来的材料。**

（八）争议的关闭与实验者回归（科学知识社会学，Collins 1985）

它到哪一步：这是本章形式层上最危险的一位。它指出：判断一个实验是否做对，要看结果对不对；判断结果对不对，又要看实验是否做对。这个回归无法靠实验本身打破，争议的关闭因此依赖实验之外的因素。

分离线：Collins 关心的是**关闭靠什么完成**（社会协商还是自然），本章关心的是**关闭之后那份残余的账目位置**。本章不主张关闭是社会性的——**即使承认关闭完全由自然事实决定，摊差依然发生**，因为终止性要求与关闭的正当性是两件事。这一点很要紧：本章的命题在科学实在论与建构论两边都成立，因此它不是那场争论的一方。

判定形式：取一个后来被公认为“关闭正确”的争议案例（自然给出了裁决），考察其关闭时的分歧技术内容是否被保留。若正确的关闭同样不留痕，则本章对而 Collins 不足以解释；若留痕与关闭的正确性相关，则应归于他那一路。

(九) B 类不确定度 (计量学, GUM)

它到哪一步：它明确承认不确定度中有一部分来自判断而非统计，并要求给出一个数。它是三家里唯一为“判断成分”设了名目的地方。

分离线：它要求把判断转化成一个数；本章说**转化成数的那个动作本身消灭了判断的内容**。一个 B 类不确定度分量告诉你“这里有一份靠判断的东西，大约这么大”，不告诉你判断的是什么、依据什么。**它是不确定度这只筐最诚实的一个角落，而它的诚实止于规模。**

(十) 稳健统计与降权 (Huber 1964)

它到哪一步：离群值不被剔除而被降权，权重是显式的、可记录的。

分离线：降权本身留痕（权重是数），但**降权的理由不留痕**——为什么这一条被判为离群、离群的方向意味着什么，不在权重里。本章要的正是权重之外的那部分。这一位与本章最接近，也最容易混：**它保留了“多少”，本章要的是“什么”。**

十三之一·划完之后：这些邻居各自看不见什么

把上面每一位再问一遍“它把什么当作给定，因此结构性地看不见什么”，句式固定：

- **计量平差**把“每条输入自带可比的不确定度”当给定，因此看不见没有不确定度的判断如何被合并——一旦承认大量判断无法自报不确定度，平差就没有输入格式。
- **信度理论**把“分歧是误差”当给定，因此看不见分歧携带的信息——一旦承认分歧携带信息，观察分数 = 真分数 + 误差这个分解就塌了。
- **安全论证**把“论证必须可关闭”当给定，因此看不见永久开放的项——一旦承认某些项不该关闭，交付就不可能。
- **德尔菲**把“收敛是目标”当给定，因此看不见有价值的收敛——一旦承认收敛可能是对的，它的整套轮次设计就没有停止条件。
- **争议关闭研究**把“关闭是终点”当给定，因此看不见关闭之后的残余——一旦承认残余持续起作用，“关闭”这个描述本身就不成立。
- **默会知识**把“没被说出来”当给定，因此看不见“说出来了又消失”的那一类——一旦承认这一类，它对失传的解释就少了一大块。

六行背后是不是同一块地？是。它可以写成一句：

一个验证程序必须能够终止；而终止意味着所有未决必须被处理掉。因此在制度层面上，“处理”与“消灭”是同一个动作。

本章把它叫做**终止性要求**。没有任何一派会去质疑它，因为互相批判的各派都站在它上面：主张更严格量化的与主张论证式判断的，主张共识程序的与批判共识程序的，全部预设了程序必须能结束。分歧只在“该用什么方式结束”。“未被任何学科辨识”的真正来源不是没有人看，是看的人都站在它上面。

本章所说的那一份，正是从这块地的缝里掉下去的：它既不属于结论，也不属于误差，它是**终止这个动作的产物**——而“终止”在所有账本上都不是一个会产生条目的事件。

十四·与本系列各章的分界

本系列近期有数篇处理相邻问题，其中一篇与本章同题，三篇与本章同学科组。它们是本章最近的邻居，必须逐篇划清，否则读者有理由怀疑这些文章在换着说法说同一件事。

《并身》（第四章）（教育学×化学×艺术学，与本章同一道题）。那一篇说的是：上一版判断在成立期间通过它所授权的执行，已经改写了复核所要面对的那个对象，而这份改变被读作对象自身的属性；它给出的读数是留底率——复核所依据的证据中，取自首次处方之前的占比。

分界：那一篇的缺栏在**对象侧**（被判断者已被改写），本章的缺栏在**判据侧的账目处理**（分歧被吸收）。两个量正交，而且可以取到四种组合：一位基线记录完整（留底率高）而会诊分歧被全部吸收（存异率零）的病人；一位急诊入院、无任何基线（留底率零）而争议巨大且写进讨论记录（存异率高）的病人。**判决性对照**：在留底率一致的一组病例中，若存异率对“复核结论被后续推翻后能否查出当初争议”没有独立解释力，则本章被那一篇吸收。

叠加效应：两者都低时，复核的结论在任何意义上都不可回溯——既没有未被自己动过的证据可以重读，也没有当初分歧的痕迹可以回查。两个机制的作用不是相加而是**相乘**：留底率 0.1 与存异率 0.1 的一次复核，其可回溯性不是 0.2，而是接近 0.01——因为回溯需要两样东西同时在场。这一点值得单列，因为它意味着两篇文章合起来给出的诊断，比任何一篇单独给出的都严重。

《阙前距》（《治未病立账学》第二章）（同三学科，处理“何谓未病”）。那一篇说的是：判定线本身在移动，而一段距离的变化中“来自对象／来自线”的构成不设字段；它给出的读数是线移占比。**分界**：那一篇处理**判据线的位置变化**，本章处理**分歧的账目去向**。两件事在时间上错开：线移发生在两次判定之间的标准修订，摊差发生在单次复核的结论形成瞬间。**判决性对照**：锁死判定标准版本（即线不动），若分歧的技术内容仍系统性消失，则本章的机制独立于线移。

《欠返》（《治未病立账学》第五章）（同三学科，处理“未病如何被识别、无症状时如何正当干预”）。那一篇处理的是未完成的返程。**分界**：那一篇的对象是**没有走完的过程**，本章的对象是**已经终止的程序所吐出的残渣**。前者缺的是完成，后者缺的是完成之后的记账。**判决性对照**：一次完整走完、无任何欠返的复核，其存异率可以为零——完成度与留痕是两件事。

《悬量》（《治未病立账学》第八章）（同三学科，处理“治未病为何实践得最少”）。那一篇说的是：阙下风险总量确定而归属未定，一切介入必须落在具体对象上，故只能落在构造出来的对象上；它给出的读数是落点自证率。**分界**：那一篇的未决在**归属**（有一份量，不知道该记到谁身上），本章的未决在**内容**（有一份分歧，知道是谁的，但内容被消化了）。二者的形式相近而位置相反：悬量是**有量无主**，摊差是**有主无量**——那位进修医生的名字在，他的判断内容不在。**判决性对照**：在归属完全明确的场合（同一位医生前后两次判断不一致，无第二人），悬量的机制不启动而摊差照样发生。

把五篇并排看，它们共有一个形式——**发生了而没有产生条目的东西**。这个形式很有生产力，而生产力太高本身就是警号。本章能给出的唯一防线，与《并身》末尾写的是同一条：若有人能证明这个形式在任意领域都能开出一样看似深刻的东西，那么这一批文章都要一起重估。本章与《并身》同题，因而排在被重估名单的最前面两位。

十五·三门学科反过来给本章加的约束

借用不是单向的。三门学科在被本章使用的同时，也各自砍掉了本章原本打算主张的一些东西。三处削减都写在这里，并写清削掉的是什么。

第一处，工程学削掉了本章最自然的那个处方。

本章原本会写下这样一句更强的话：“**复核中出现的分歧，应当全部保留。**”这句话很顺，而且看上去只是一个记账要求，成本很低。

工程学不允许它成立。工程学是三家里唯一真的做过这件事的一家——它有开放项清单制度，几十年了。而它的实际经验是：**开放项清单会膨胀到无法使用。**一份大型系统的安全评审可以留下数以千计的开放项，最终结果是没有人读，等于没留；更糟的是，清单一旦过长，关闭动作会退化为批量走形式，**留痕制度本身变成了一台新的吸收装置。**合理可行降低原则之所以存在，正是承认必须关闭一些东西才能交付。

因此本章的主张必须削弱为：**不是全部保留，而是关闭时必须留下关闭理由的技术内容；并且必须规定回读的触发条件，否则留痕等于不留。**这是一处实质削弱：它把“保留分歧”这个听起来无成本的要求，换成了一个有筛选、有格式、有回读机制的制度设计，成本大得多，而且**筛选本身又是一次会产生摊差的判断**（见第十七节的反噬）。

第二处，物理学动了本章的适用范围与方法主张。

本章原本会写下这样一句更强的话：“**复核在结构上是一次全局平差，因此中医的复核也应当按平差来组织。**”这句话有气势，而且顺着第四节材料很容易滑出去。

物理学不允许它成立。平差之所以能工作，靠的是一个中医没有的条件：**每一条输入自带一个可比较的不确定度。**一次瓦特天平测量自报一个数，一次硅球计数自报一个数，两个数可以在同一个方程组里被加权。而“主治判为肝郁脾虚”这条输入不带不确定度，也无法被要求带——把它写成“我有 70% 的把握”是一个伪精确，那个数字没有任何可比性。

因此本章必须收窄为：**复核在结构上是配权，但在缺乏可比不确定度的场合，能做的只有登记，不是配权。**这是一处很重的削减：它把本章从一个方法主张退回到一个记账主张。本章因此不主张任何形式的“辨证平差”，也不主张给辨证结论标注置信度——后者会立刻成为第四只筐。

第三处，心理学动了本章的价值排序。

本章原本会写下这样一句更强的话：“**吸收分歧是一种损失。**”

心理学不允许它成立。心理测量几十年的证据说明，共识程序确实在做工作：结构化会谈的诊断预测效度高于自由会谈，多评判者合并的判断优于单个评

判者，训练能真实提高一致性且这种一致性并非全是趋同假象。吸收不只是掩盖，它把一堆不可用的判断加工成了一个可用的结论。

因此本章必须放弃“吸收=损失”的表述，改为：吸收是必要的加工，问题只在于加工的残渣不留账。这一处改动最伤气势——它使本章无法站在批判标准化的位置上，只能站在一个技术性得多、也小得多的位置上：加一栏。但它也使本章的主张变得可能被接受，因为它不要求任何人放弃任何已经证明有效的做法。

十六·不适用的边界

本章的命题在下列情形不适用，援引者应当自己先排除：

一、有外在锚可及的场合。若存在一个不参与判断生成过程的金标准（病理诊断之于影像判读、基准器之于日常计量），复核可以是真正的裁决，本章的机制退居次要。

二、分歧本来就极少的场合。存异率的分母为零时这个读数不适用。多数常规、成熟、低歧义的判断落在第②格，无害，不需要本章。

三、单次判断、不设复核的场合。没有第二次判断，就没有本章所说的问題。

四、需要立即决断的急症。分歧留痕不得推迟处置。急诊场合应当先出结论，留痕补记——而“补记”这个动作本身会损失内容，这是必须承认的代价，不是可以靠制度设计消除的东西。

五、本章交出去的第一个洞：分母不可得。存异率的分母是“复核中出现过的分歧条目总数”，而在没有过程记录的场合，这个分母恰恰是被吸收掉的那样东西——要测量摊差，需要一份摊差尚未发生时的记录。这是一个真实的循环。本章的处置是只在有过程记录的场合计算它（会诊记录、德尔菲轮次表、评审记录、平差输入清单），并承认在没有过程记录的场合，摊差在原理上不可测。这不是一个可以靠更好的设计解决的困难。

六、本章交出去的第二个洞：什么算“一条分歧”。两位医生一个说肝郁脾虚一个说湿热中阻，这是一条分歧还是三条（病位、病性、主次）？粒度不同，存异率可以差一个数量级。本章只能要求任何比较研究预先固定粒度规则，而无法给出一个有原则的划分。这一条会显著限制存异率的跨机构可比性。

七、本章交出去的第三个洞，也是最大的一个：留痕的价值依赖于有人回读，而回读的场合极少。第十节场合二说明了这一点——物理学的输入清单公开了几十年，下游使用者仍然只取那个数。一个没有人读的字段，与一个不存在的字段，在效果上难以区分。本章第十二节第五条试图用回读触发条件来补，但那只是把问题推后一步：谁来保证触发条件被执行？本章没有答案。

十七·证伪条件、样本纪律、伦理边界，与一条写死日期的赌注

十七之一·最强的竞争解释

在给出证伪条件之前，先把那三句最难反驳的话摆在明处：

“这其实不就是共识程序在正常工作吗？” “这其实不就是集体思维、有人不敢说话吗？” “这其实不就是有些东西本来就说不清楚吗？”

一条证伪条件如果不能排除这三句，它只是在验证“本章描述的现象存在”，那不算数。以下条件按能否排除竞争解释来设计。

条件一（核心，机制证伪）：控制评审人数、评审时长、参与者资历构成之后，若“程序的一致性目标强度”（是否设共识度阈值、阈值高低、是否有资历裁决人、是否要求给出唯一结论、是否按信度删条目）与“分歧技术内容在归档文件中的可检索率”之间**不存在剂量-反应关系**，则本机制为伪——届时留痕的缺失应归因于记录习惯而非程序结构。

条件二（排除集体思维）：在完全匿名、无社交互动、由算法自动合并的评判场合，若分歧的技术内容同样不可检索，则本判断对；若匿名化之后留痕显著改善，则应归于集体思维。

条件三（排除默会知识）：只统计曾以文字形式出现在过程记录中的分歧条目。若这部分同样大比例不可检索，则不是“说不出来”，是“说了被吃掉”。这一条从设计上把默会知识完全排除在外。

条件四（低成本、可立刻实做，数据已经存在）：取任一部已颁布的证候诊断标准，调出其德尔菲各轮原始打分与理由记录；把被删除条目与保留条目一起，在一个独立的病例集上考察其对治疗选择或临床结局的辨别力。**若被删除条目**

的辨别力不高于保留条目，本章的核心主张失去它最重要的一处支撑。所需数据在多数标准制定单位的档案里已经存在，一名研究生数月可以完成。

条件五（正向读数，顺序预测）：在引入“未采纳意见须留技术理由+回读触发条件”制度的机构，**复核结论被后续推翻后“能查出当初争议”的比例上升，应当出现在制度实施后 2-三个复核周期，而不是与制度宣贯同步出现。**若改善与宣贯同步、且在第一份存异记录被实际回读之前就已完成，则改善应归于宣贯的注意力效应而非本机制。**顺序比相关难被巧合满足得多。**

条件六（分层引用与分层证伪）：本章的主张分三层，可分层引用、分层推翻。第一层是“摊差”这个构念（最强、最易倒）；第二层是那张两轴四格的辨别表（作为分析工具，即使第一层被推翻，第③格与第④格的区分仍可用）；第三层是一条不依赖前两层的经验主张——**“证候标准制定过程中被删除的条目，其临床辨别力平均不低于保留条目”**。第三层最稳、最便宜、最容易检验；若连第三层都被推翻（即被删条目确实辨别力更低），本章全盘作废。

十七之二·样本纪律

上述条件涉及跨案例比较的，**样本不得少于六个同质单元，且在关键混淆变量上同质。**可用单元：同一医院不同科室的会诊流程、同一学会不同工作组制定的标准、同一企业不同项目的安全评审、同一期刊不同栏目的评审记录。**严禁拿两个国家或两个行业的对比充数——**两边在记录文化、法规要求、信息化程度上处处不同，所谓“控制掉某变量”在方法上站不住；这类比较只能作启发式例证，不能充当证伪检验。

十七之三·伦理与证据边界

存异率是一个比率，一个比率一旦提出，几乎必然被拿去考核人。以下三条必须写死，缺一条本章不宜被引用：

一、**这个读数指的是位置，不是人。**存异率低说明“这套程序吃掉分歧不留痕”，不说明参与者水平差或不诚实。尤其不得用于评价提出不同意见的那位——**那会直接摧毁本章想要保护的东西。**

二、**不得为了把这个数做好看而制造分歧或延迟决断。**任何以“提高存异率”为目的的人为争议、拖延结论、保留无意义的反对意见，都是不可接受的。急症场合先出结论、后补留痕。

三、留痕内容不得用于事后追责。这是本章全部伦理条款里最要紧的一条：若“未采纳意见”栏可以被用来追究采纳者的责任，这一栏会在三个月内变成空栏或套话栏。制度设计上必须明确该栏的用途限于回溯与学习，并给出相应的保护。没有这一条，第十二节的全部动作都会失败。

十七之四·反噬预言

本命题一旦被制度采用，最省事的达标方式不是留痕，而是把“未采纳意见”栏填成套话——“经讨论，其他意见依据不足，未采纳”。这一句满足字段要求，不含任何技术内容，成本为零。更糟的一种是它诱发分歧的制造：为了让这一栏好看，安排一位固定的“提反对意见者”，提一些安全的、无关紧要的反对。

而这个反噬有一个自指的形状，使它格外难解：“哪些分歧值得留痕”本身就是一次会产生分歧的判断，而那一次分歧同样会被吸收。也就是说，本章的处方在执行时会立刻生成一份新的、更高阶的摊差，而针对它的处方又会生成下一份。

我看不到出口。唯一能想到的部分缓解是第十二节第六条——署名，使留痕对留痕者本人有价值，从而不完全依赖行政强制。但署名同时提高了留痕的社会成本，两边不能同时优化。把这一条写在这里，是因为一个看得到出口的反噬不是反噬，只是待办事项。

十七之五·一条写死日期的赌注

到二〇三三年十二月三十一日，至少有一部由国家级中医药行业组织颁布的病历书写规范、会诊制度规范或证候标准制定技术要求中，出现如下强制条款：会诊、疑难病例讨论或标准制定的归档文件须单列“未采纳意见”，内容须包含分歧所在条目、不同判断的内容、其临床依据三项。

什么不算命中，四条写死：

1. 仅出现在指南讨论稿、专家建议、学术论文或教材中，未进入颁布文本的，不算；
2. 仅要求“记录讨论过程”“完善讨论记录”而无三项内容要求的，不算——这一条尤其要紧，因为“记录讨论过程”这句话在现行规范里已经有了，而它正是本章所指的空转；

3. 仅要求设立字段而未规定内容最低要求、或允许以“依据不足未采纳”一类套话填写的，不算；
4. 由本章作者或其所属机构参与起草的文本，不算。

十八·本章自己是不是它所描述的那样东西的标本

按本章的判据，本章自己落在哪一格？

先算本章的存异率。本章在成文过程中处理过三门学科之间的多处冲突，其中相当一部分**没有被解决，而是被本章的框架吸收了**。举一处最明显的：第二节说物理学与心理学对“共识能否作为算数的依据”给出不相容的归属——这个冲突在本章里没有被裁决，它被转化成了“两家各锁一个终点”这个说法，而“各锁一个终点”正是一只筐：**它把一场无法调解的分歧，转化成了一个整齐的分工描述**。

本章对这类吸收留了多少痕？留了一些——第十五节逐条写了三处削减，第十六节列了三个洞。这些是本章的存异栏。但它同样有大量未留痕的：三家在本章里的形象是被本章整理过的，各家内部的争议、各家对本章这种读法可能有的反对，本章没有逐条列出。**估算本章的存异率，大约在 0.3 上下——高于多数论文，远低于 1。**

这不是一句谦辞，它有三条具体后果：

第一，本章对三家的描述都是被这个框架整理过的。例如第六节把“结合临床，随证治之”读成第三只筐——这句话确实可以这样读，但它是否本来就该这样读，还是被本章的框架推到了这个位置上，本章自己分不出来。要分出来，需要一份在本章框架成形之前就已固定的解读记录，而本章没有留。**这与本章的论题姊妹篇《并身》所指认的机制是同一个，两篇文章互为对方的标本。**

第二，本章最不该被引用的方式，是被当作已经证成的东西引用。本章能提供的只有第十七节条件四那一条——一个便宜的、任何人可以独立做出来的清点。**在那个清点被独立完成之前，本章的地位是一个待检验的读数，不是一个结论。**

第三，本章正落在自己所指认的第④格里，而且落得比多数文章更深。因为它是一篇要求给出唯一结论的学术论文——学术论文这个体裁本身就是一台吸收装置：它要求有一条承重命题、要求论证收敛、要求结论明确，而这三条要求每

一条都在吃掉写作过程中出现的分歧。**本章越是写得干净、结构越是整齐、论证越是收敛，它的存异率就越低。**这正是第八节签名一所描述的形状：质量指标的改善与摊差的增加是同一个动作的两面。

对此本章只能做三件事，而且都已经做了：第十五节写清三处削减、第十六节交出三个洞、第十七节把最便宜的那条检验单独列出并写明它不需要本章作者参与。一篇落在第④格里的文章，唯一的自救方式是**把能推翻自己的那件事做得足够便宜。**

最后一句关于这一整批文章。本系列近年数篇——包括本章与它的同题姊妹篇——共有一个形式：找出一样“发生了而没有产生条目的东西”。这个形式的生产力太高，本身就是警号：**一把在任何门上都拧得动的钥匙，通常不是钥匙，是撬棍。**本章能给出的唯一防线是第十四节末尾那句话，而本章比其他几篇更没有资格例外——因为它自己刚刚论证过，一个吸收装置在运行得最好的时候，正是它最看不出问题的时候。

十九·伦理与证据边界声明

本章提出的存异率是一个结构读数，不是绩效指标。禁止用于个人考核与晋升评价；禁止用于评价提出不同意见者；禁止以提高该读数为目的制造争议或延迟必要的临床决断；留痕内容不得用于事后追责，其用途应限于回溯与学习；据该读数作出的任何不利安排，须公开编码规则并提供复核通道。本章所举临床情形均为类型化描述，不涉及任何具体病例、机构或个人。本章不构成任何诊疗建议，也不构成对任何已颁布标准的技术评价。

注释

1. 本章所称「证」指中医辨证论治体系中的证候判断；所称“复核”指对已确立之证的再次判断，含同一诊次内多人评判与不同诊次的前后评判两种情形。
2. 第二节对三门学科验证范式的描述取其主流形态，不代表该学科内部无争议；三家之间的冲突为本章所整理，各家并未如此互相表述。
3. 第四节对全局平差与扩展因子的描述取其操作结构，具体算法、判据与历次调整详见相关方法说明文件；本章不评价该做法在计量学内部的适当性。

4. 第九节存异率的定义中，“一条分歧”的粒度会显著影响读数，见第十六节第六点；任何比较研究须预先固定粒度规则并公开。
5. 第十节场合五所述两种法域安排，本章只取其形式对照，不涉及任何司法制度评价。
6. 第十七节赌注的判定以颁布文本的正式发布日为准。

第六章 填判

——一个证被带出它的适用范围时为什么不会报错——论“不适用”为什么不是一个可以填的结论，以及它该被填在哪里

提要 一个证被确立，被复核，被写进病历，被带到下一位医生、下一家医院、下一项研究里去。这条链上有一个环节从未被检查：**这套判据在这个人身上到底适不适用**。现行的一切验证体系只有两种输出——判定成立，或判定不成立；**没有第三种输出叫“此处无从判定”**。于是所有本该落在这第三种位置上的情形，被强行填成前两格之一，而填出来的结论与正常结论在形式上完全一样，没有任何标记可以区分。本章把它命名为**填判**——一个本该空着的位置，被系统用它唯一有的那几个输出填上了。本章的承重命题是：一个证的复核，不是看它现在的表现不符合这一型的标记，不是看作出这个判断的程序正不正当，也不是看它在规定范围内是否被验证过，而是**先判定“这个判定在此处适不适用”，再判定它对不对**；而第一道判定在现行体系里没有输出格式，因此它从未被作出。材料取自细胞学的类型注释、工程学的型式试验与适用范围、管理学的程序正当性三处成熟范式的互相顶撞；推翻它们共有前提的那件证据来自细胞学自己——一个细胞被判为哪一型，取决于它与哪一批细胞一起被聚类、在什么分辨率下被切分；换一个参照集合，判定就变，而算法必然给每一个细胞分配一个标签。本章给出一个可在四个领域并排清点的读数——**弃判率**：一段时期的全部判定输出中，被记为“不适用／无从判定”且带有下游处置路径的次数占比；并给出一句不含情态词的问话：**上一次复核的结论表上，“无从判定”有没有一个可填的格？过去一年被填过几次？**结论是：一个判定的可信度，第一位不在它判得准不准，在这套体系允许不允许它说“这里不该由我来判”。

一·一个证被带走的时候，什么没有跟着走

先摆一个每天都在发生、而没有任何一方觉得需要解释的场景。

一位病人在南方某院就诊，医生辨为湿热蕴脾，开方，病历上写下这四个字。三个月后他到北方另一家医院，医生调阅病历，看到“湿热蕴脾”，在此基础上继续辨证。

“湿热蕴脾”这四个字被带走了。什么没有跟着走？

跟着走的是一个结论。没跟着走的是：这位医生每天看的是哪一类病人，他手上“湿热”这个概念是在哪一批人身上校准出来的，那个地方的气候与饮食把他的判别阈值推到了哪个位置，他所受的师承把哪几条舌脉抬成了主症。**这些东**

西不是他的偏差，它们是他作出这个判定的凭据。判定被带走了，凭据留在了原地。

有人会说：那正是中医不够标准化的表现，把标准统一了就好了。

这句话值得停下来看。它假定存在一个可以脱离一切参照条件而独立成立的判定——“**湿热蕴脾**”是这位病人身上的一个客观事实，谁来看都该看到它。而本章要说的是，这个假定在**任何**判定体系里都不成立，中医只是没有掩盖它。三门与中医毫不相干的学科，各自在自己的领域里撞上同一件事，各自建了一套装置去对付它，而三套装置互不兼容。

更要紧的是第二个场景，它比第一个隐蔽得多。

一位病人来看诊，他的情况不在这位医生所熟悉的谱系里——一个罕见的体质、一种从未见过的组合、一个长期服用多种西药后被改造过的身体。医生辨证，写下一个证名。

这个证名与前一个场景里那个证名，在病历上长得一模一样。

没有任何标记区分“我很有把握”与“我按最像的一个填了”。没有任何字段可以写“这套判据在这个人身上恐怕不适用”。病案首页的诊断栏不允许空着——这是硬性规定，填写率是要被考核的。于是那个位置**必须**被填上一个东西，而系统能提供的东西只有若干个证名。

这不是中医的毛病。请看三处完全同构的情形：

- **一个细胞被送进聚类算法。**算法必然给它一个类型标签——k 均值必然分配，即使是更精细的社区发现算法，也极少输出“我不知道”。一个处在分化中间态、不属于任何已知类型的细胞，会被判给离它最近的那一型。**输出集里没有“不属于任何一型”。**
- **一台设备被开到型式试验从未覆盖过的工况点。**它照样能开机，照样运行，没有任何提示说“此处未经鉴定”。合格证上写着适用范围，而使用端没有任何机制去读它。
- **一套合规审查程序被用在它未被设计针对的情形上。**它照样能走全部环节，照样给出“符合”或“不符合”的结论。**没有一栏叫“本程序不适用于此情形”。**

三处的形状完全一样：**验证体系的输出集里没有“不适用”，于是所有不适用的情形被填成了适用。**

于是问题可以写成一句话：

一个证如何被确立并被复核，才算数？

这个问句里有一个被默认的顺序：先有一个证，然后检查它对不对。本章要说的是，在“对不对”之前还有一问，而这一问从来没有被提出过——不是因为没人想到，是因为**提出它之后没有地方填答案**。一个没有输出格式的问题，等于没有被问。

二·三种成熟的验证范式，与它们各自锁死的那个终点

三门学科都要回答同一个问题：一个判定凭什么算数、又凭什么被推翻。三门学科给出的答案不能并存。

二之一·第一家：细胞学——算数落在它现在表达什么

一个细胞属于哪一型，判据是它当下表达的那一组标记。

这一家在过去二十年经历了一次彻底的技术转向：从形态与位置转向分子标记，再转向单细胞层面的全转录组。今天判定一个细胞的类型，标准做法是把成千上万个细胞的表达谱放在一起降维、聚类，让相似的聚成一团，再给每一团指认一个类型名。判据完全落在**当下测到的那一组数**上：不问它从哪儿来，不问它将来会变成什么，不问它经历过什么。

这一家的终点是当下的表达。复核动作是**重新测一遍、重新聚一遍**：换一批样本，看同样的类型能不能被重新找出来。

它能成立，靠一个不写在方法学里的前提：每一个细胞都属于某一型。整套流程的输出格式就是“每个细胞一个标签”——没有“不属于任何一型”这个输出。质控环节会剔除一部分细胞（双细胞、死细胞、低质量），但**剔除是删除，不是标注**：被剔除的细胞从数据集里消失，不会以“未知”的身份保留在最终图谱里。

二之二·第二家：工程学——算数落在规定的使用范围之内

一个产品算合格，是因为它在一个被明确划定的使用包线内被逐点验证过。

型式试验的做法是：把使用条件写成一个多维的范围——温度从多少到多少、湿度、振动谱、电压波动、负载区间；然后在这个范围的角点与关键点上逐

一施加考验，通过全部考验才发证。合格证上写明适用范围，超出这个范围，证书不担保。

这一家的终点是那条包线。复核动作是**周期性再鉴定与包线扩展试验**：设备用久了重新考一遍；要用在新工况上，先做扩展试验把包线推出去。

这一家最值得注意的是：它是三家里唯一显式记录了作用域的一家。合格证上真的写着“适用于 -20°C 至 $+55^{\circ}\text{C}$ ”。这一点后面要反复用。

而它能成立，靠一个前提：包线内部是可内插的。考了四个角点就认为整个矩形合格。工程学自己知道这个假定有反例——包线内部可能藏着谐振点、临界工况、参数组合效应，而角点试验恰好跳过它们。

二之三·第三家：管理学——算数落在作出判断的那个程序上

一个决策算数，不是因为结果好，而是因为作出它的程序是正当的。

这一家有一条被反复强调、也被反复违反的原则：好决策不等于好结果。结果受运气影响，用结果反推决策质量是一种系统性偏误。因此它的复核对象是**程序**：是否掌握了应有的信息、是否考虑了备选方案、是否有回避与制衡、是否留下了可查的记录、是否由有权限的人在有限范围内作出。

这一家的终点是那条程序。复核动作是**合规审查与决策复盘**：把程序重走一遍，看每一步是否被遵守。

它能成立，靠一个前提：程序总是适用的。一套审查程序被启动之后，它必定给出“符合”或“不符合”。而“本程序不适用于此情形”这个结论，在绝大多数合规体系里不是一个可选项——它意味着审查无法完成，而审查无法完成在管理上等同于失职。

二之四·三家在打架，而且打得没法调解

细胞学与管理学：细胞学的整个转向就是为了**只看结果不问过程**——不问这个细胞怎么来的，只看它现在表达什么。管理学的核心原则恰恰相反：**结果不能作为判断依据，只有过程可以**。两家对“当下的表现能否作为算数的依据”给出不相容的归属。若细胞学对，管理学对结果偏误的整套批判是多虑；若管理学对，细胞学把身份完全押在一个瞬时快照上是极其危险的。这不能靠判谁对化解——两家不在同一个语域里作答。

细胞学与工程学：这一对是最硬的一对。工程学每天在做的事是**把对象拿到包线边缘去考**：高低温循环、振动、加速老化、过载，故意逼到极限看它还成不成立。细胞学明确把这件事诊断为病理——**把细胞拿到非生理条件下，它的身份就变了**。培养基一换、贴壁条件一变、传代多几次，表型就漂移：原代肝细胞离体培养一天之内就丢掉大部分肝特异功能；上皮细胞在特定条件下整体转成间质表型。细胞学因此有一整套“最小扰动”的规矩：原位、活体、少传代、尽量在原生境里看。反过来，细胞学的日课——在原生境里最小扰动地观察——在工程学是**不充分鉴定**：只在标称工况下测过的东西不算合格。一门学问的日课，正是另一门学问诊断为病根的那件事。

管理学与工程学：工程学要求**穷举包线并逐点验证**——包线内的关键点都要考到，考不到就不能声明。管理学认为这既不可能也不必要：判断永远在信息不完全的条件下作出，程序正当性的全部意义正是在**无法穷举时仍然能给出一个可辩护的判断**。若工程学对，管理学的决策质量评价是自欺；若管理学对，工程学的包线鉴定是在假装穷举——而包线本身是人划的，划在哪里就是一次判断。不能同真。

三家各锁一个终点：**当下的表达、那条包线、那条程序**。三个终点互不隶属。而正是在它们争夺终点的时候，三家脚下踩着同一块地。

三·三家共同假定的那件事：判定是可以被携带的

三家争的是一个判定的效力该由哪一个终点来结算。争得很凶，而争的形式本身暴露了一条谁也没说出口的假定：

一个判定一旦被确立并复核通过，它的效力可以脱离当初的验证条件而被携带。

说得更直白些：**证书是可以带着走的**。一个细胞被标注为某型，这个标签就跟着这个细胞进入下游的所有分析；一个部件通过鉴定，合格证跟着它进入任何一台整机；一次决策通过合规审查，“已审查”这个状态跟着它进入所有后续追溯。判定与产生它的条件在这一刻分离，而分离之后，判定单独流通。

把这句话念给三家听，三家都会说“这还用说吗”，而且三家都会立刻指出自己有装置保证它——

·细胞学说：所以我们做批次校正、做整合、做参照图谱映射。

- 工程学说：所以我们在证书上写明适用范围。
- 管理学说：所以我们规定了程序的适用情形与权限边界。

三套装置，做的是同一件事：让判定在被携带时，把它的条件也带上一部分。而三套装置有一个共同的形状——它们全部装在生产端，没有一件装在使用端。批次校正是做图谱的人做的；适用范围是发证的人写的；权限边界是定程序的人定的。读到它们、并在超出时停下来，是使用者的事，而没有任何一套装置在使用端设了检查点。

现在把这一条搬到辩证论治上。它当场断裂，而且断的方式与本系列问题的另外两篇所指出的都不同。

- “**湿热蕴脾**”这个标签是可携带的——它在病历上、在转诊单上、在研究入组表上、在医保编码里流通得非常顺畅。
- 产生它的那一套参照条件完全不可携带——这位医生见过的病人分布、他的判别阈值、这个地域与季节、这一批同行的用语习惯。这些东西没有任何字段可以承载，也没有任何人认为它们需要被承载。
- 更要紧的是：没有任何一个环节会在标签被超范围使用时报错。下一位医生拿到“湿热蕴脾”，不会收到任何提示说“这个判定是在一个与你完全不同的参照集里作出的”。

有人会说：那就把标准统一了，让所有医生用同一套参照。

这条路正是中医现代化过去几十年在走的路，而本章第四节要说明：它在细胞学那里已经被走过一遍，并且撞上了一堵墙——不是因为做得不够好，是因为“统一参照”这件事在原理上做不到。细胞学自己有这方面最硬的证据，而那份证据在它自己的领域里回答的是另一个问题。

四·细胞学自己的一件材料：一个细胞是哪一型，取决于它和谁一起被聚类

要推翻一条三家共同踩着的假定，从外面搬理由来没有用——那只是加入争论，成为第四家，三家可以各自退回阵地。要让三家退不回去，推翻它的那件材料必须是三家之一自己已经掌握、已经公开、只是从来不往这个方向用的事实。

这件材料在细胞学手里，而且是它每天在做的一件事。

判定一个细胞的类型，标准流程是：把一批细胞的表达谱放进同一个空间，降维，构建近邻图，用社区发现算法切分成若干团，再给每一团指认一个类型名。这条流程有两个参数是人定的：**这一批细胞是哪一批（参照集），以及切分的粗细（分辨率）。**

细胞学自己非常清楚这两个参数的后果，而且写在每一篇方法学讨论里：

第一，同一个细胞放进不同的参照集，会被判为不同的类型。把一群 T 细胞单独拿去聚类，会切出若干个亚型；把同一群 T 细胞混进全血的所有细胞里一起聚类，那些亚型会合并成一团，全部被标为“T 细胞”。**没有任何一个判定是错的——它们回答的是不同的问题。而这个细胞本身什么也没变。**

第二，分辨率参数直接决定有几种类型。调高一点，一团裂成三团，于是“发现”了两个新亚型；调低一点，三团并成一团，那两个亚型消失。这个参数没有客观的最优值；实践中的选择依据是“看起来合理”“与已知标记吻合”“下游分析可解释”。

第三，也是最要紧的一件——算法必然给每一个细胞分配一个标签。近邻图上的每一个点都会落进某个社区。一个处在两团之间、不属于任何一型的中间态细胞，会被判给它偶然更靠近的那一边。**输出格式里没有“不属于任何一型”这一项。**质控会剔除一部分细胞，但剔除是从数据集里删除，不是在图谱上保留一个“未知”的记录——**剔除之后，那个细胞连同它所提出的那个问题一起消失了。**

请看清楚这三条合起来意味着什么：

一个细胞被判为哪一型，不是这个细胞的属性，是这个细胞在一个特定参照集中、在一个特定切分粗细下的位置的属性。而这个判定被携带进下游分析时，参照集与分辨率都不跟着走。

细胞学人当然知道这一条，而且发展出了一整套对策——批次效应校正、参照图谱映射、聚类稳定性评估、多分辨率注释。但这整套讨论回答的是“**如何让注释更稳健、更可重复**”。它从来不被拿来回答“**一个判定的效力能不能被携带**”。这正是它能推翻共有前提的原因：在它自己的领域里，它回答的是一个数据分析质量问题。

把它读成一条关于验证结构的证据，得到的是：

判定不是被携带的，是被重新生成的；而每一次重新生成都依赖一个当场的参照集。判定看上去可携带，只是因为承载它的那个符号可携带。

这一句一旦成立，共有前提垮掉：符号在流通，判定没有在流通。而三家的生产端装置——批次校正、适用范围、权限边界——全部是在设法让符号更可信，没有一件是在使用端问一句“我这里的参照集与产生它的那个一样吗”。

第二件材料在工程学手里，可以顺带一提，因为它从另一头证明同一件事：**包线角点试验的内插假定**。考了矩形的四个角就认为整个矩形合格，而工程学自己知道包线内部可能藏着谐振点与参数组合效应。这意味着：“**在适用范围内**”这句话本身也是一次判定，而它同样没有被单独验证过。适用范围是三家里唯一被显式写下来的作用域，而即使写下来了，它内部仍有大片区域从未被考过——那些区域被记成了什么？**被记成了合格。因为没有“未验证”这一栏。**

第三件材料在管理学手里：**有限理性与满意解**。管理学自己承认，判断永远在信息不完全的条件下作出，穷举不可能，人只能找到“足够好”的方案。**它自己知道判断是有限的，而它的记账是无限的**——绩效与合规的输出仍然是二值的：达标／未达标、符合／不符合。承认判断有限，却不给“超出我判断范围”留一个输出格式，这两件事同时存在了几十年而没有被并排看过。

到这里可以把话说到底：三家看不见这一层，**不是因为它们疏忽，是因为它们各自的核心问题恰好使这一层成为不必要的**。细胞学的核心问题是“这一批细胞由哪些类型组成”——它要的是一张完整的图谱，图谱上不能有空白，**每个细胞都有标签正是它的交付要求**。工程学的核心问题是“这东西能不能用”——它要的是一张证，证上必须有一个明确的适用范围，**范围内一律合格正是证的意义**。管理学的核心问题是“这个判断有没有人负责”——它要的是一次可追责的结论，而“不适用”这个输出把责任推回给了提问者，等于没有结论。三家的盲区，是三家各自交付要求的影子。

五·填判：一个本该空着的位置被填上了

现在可以给那样东西命名了。

任何一次判定，都发生在一个参照条件之中：一批被比较的对象、一条被划出的界线、一套被默认适用的程序。当被判定者落在这个条件之外时，正确的输

出应当是“此处无从判定”。而三家的输出集里都没有这一项，于是那个位置被系统用它唯一有的那几个输出填上了。

本章把它叫做**填判**：

填判——被判定者已落在判据的参照条件之外，而验证体系的输出集中不含“不适用／无从判定”，因此该位置被填以一个形式上完全正常的判定；填出来的结论与在参照条件之内作出的结论没有任何可区分的标记。

“填”取的是两重意思：填表的填——那一栏不许空着；以及填坑的填——一个本该显示为空缺的位置被抹平了。

要看清它是一样独立的东西，最快的办法是把它与四个相邻的东西分开：

它不是误判。误判有一个隐含前提：存在一个正确答案，而你给错了。填判发生在**没有正确答案**的场合——不是“这个人其实是脾虚而你判成了湿热”，是“在这个人身上，湿热与脾虚这套区分本身可能不构成一个有意义的问题”。把填判记为误判，就是把一个第三态压回二值，也就是本章要指认的那次填充本身。

它不是不确定。不确定是在一个适用的判据下，证据不足以定夺——它有明确的补救路径：多查一项、多观察几天、再请一位医生。填判无法用这些办法补救，因为再多的证据也不会告诉你“这套判据不该被应用”。这一条分界后面要用来构造一条硬的证伪条件。

它不是“未特指”。疾病编码体系里有“未特指”这一类：知道是这一类病，但没有细分。**它是一个阳性归属**，可以入组、可以结算、可以统计。填判要的是**归属的悬置**——不是“是这一类但没分清”，是“不知道能不能算这一类”。两者在账上完全不同，第十三节会正面处理这一位，因为它是最容易被拿来说“我们已经有了这个字段了”的一位。

它不是“待查”。临床有“发热待查”这样的写法，它是**时间性的悬置**：现在不知道，检查推进就会知道。填判是**结构性的不适用**：这套判据在这个对象上不成立，而时间不会改变这一点。待查会闭合，填判不会。

填判的三条性质，本章后面全部要用：

1. **它是被正确运行的体系所产生的，而且是被要求产生的。**不需要任何人偷懒或违规。诊断栏不许空着，这是明文规定；填写率是被考核的指标；病案质控会因为空栏而扣分。**填判是合规的产物，不是违规的产物。**

2. 它与正常判定在形式上不可区分。同一个证名，写在同一栏，用同一套编码，进入同一个统计口径。没有任何下游程序能把两者分开——这是它与所有已知失效模式最不同的一点。

3. 它不可归咎，因而无人有动机记录它。那位医生做了他能做的最好的事：在没有“不适用”这个选项的情况下，选了最像的一个。他没有做错任何事。记录制度是被责任驱动的，于是没有人为它开一栏。

三条合起来是一个少见的形状：一样确实发生了、可能很频繁、对下游有决定性影响、而且没有任何一方做错事的东西，因此它在所有账本上都不设字段——不是记录得少，是输出集里根本没有这一项。

六·三家各自把“不适用”填到哪里去了

填判不会凭空发生，它总是被填成某一个具体的输出。三家各自的填法不同，看清这三种填法，是本章最实用的一节。

第一种填法：填给最近的那一类（细胞学）。

聚类算法把中间态细胞判给它偶然更靠近的那一团。这种填法的特点是它有方向性：填出来的结果不是随机的，而是系统性地偏向参照集里样本量最大的那一类——因为大类的引力更强。于是一个原本处在边缘的对象，被并进了主流类别，而它作为“边缘”这一事实被消灭了。

在中医里对应的是什么？对应的是**常见证型的膨胀**。一位医生手上最熟的那三五个证型，会吸收掉大量本该落在别处或落在无处的病人。这不是水平问题——任何判定体系都会向它的密集区塌缩。而临床研究的证型分布数据，正是由这样的判定累积出来的，于是“某证型占62%”这个数字里，混着不知多少填判。

第二种填法：填成“在范围内”（工程学）。

包线内的未考点被默认为合格。这种填法的特点是它**偏向肯定**：不确定的一律算通过，因为不通过需要证据，而通过是默认状态。**未验证与已验证在证书上长得一样。**

在中医里对应的是什么？对应的是**证候标准的适用人群被默认为全体**。一部标准在若干家医院、若干百例病人身上做出来，颁布之后适用于所有人。**标准文**

本里几乎从不写“本标准的制定样本来自哪个地域、哪个年龄段、哪个用药背景，超出这个范围未经验证”。这一栏在工程学是强制的，在这里是空的。

第三种填法：填成“程序已完成”（管理学）。

一套不适用的程序被走完，输出“符合”。这种填法的特点是它有责任人的签字：有人签了字，说明有人为这个结论负责。而他负责的是程序被完整执行，不是程序适用于此。签字记录的是完成，不是适用。

在中医里对应的是什么？对应的是会诊与病案质控的通过。一次会诊按流程走完，结论写下，签字，归档。质控检查的是格式完整、要素齐全、逻辑自治——没有任何一条质控标准问“这套辨证体系在这位病人身上适用吗”，因为这个问题一旦被问出来，质控就无法完成。

三种填法并排看：

	填成什么	偏向	有无责任人
填给最近的一类	参照集中的密集类别	向主流塌缩	无
填成“在范围内”	默认合格	向肯定偏	无
填成“程序已完成”	符合	向完成偏	有，但他负责的是完成不是适用

三种填法各有各的偏向，而没有一种是随机的。这一点极其要紧：如果填判是随机噪声，它会在大样本里被平均掉；而它有系统性方向，它会在大样本里被放大，并且被读成一个真实的分布特征。

这也就是为什么本章要打的不是某一种填法，是“输出集里没有第三项”这件事本身。

七·两根轴与四格

收成一张辨别表，需要两根轴。

第一根轴：被判定者落在判据的参照条件之内还是之外。

第二根轴：这套体系的输出集里有没有“不适用／无从判定”这一项。

	输出集里有“不适用”	输出集里没有“不适用”
落在参照条件之内	① 正常判定——那一项用不上，但它在。	② 侥幸正常——碰巧都在域内，所以从未出事。多数成熟、稳定、同质的判定场合落

		在这里，无害。
落在参照条件之外	③ 弃判——判为不适用，并转出。法院裁定本院无管辖权并移送，是这一格最成熟的状态。	④ 靶格

第④格：被判定者已在参照条件之外，而体系没有“不适用”这个输出，于是给出一个形式上完全正常的结论。

八·靶格：所有完备性指标都在改善的那一格

第④格不是一个坏格子。它是一个看上去最完整的格子，而且是被整个规范化事业主动追求的那个格子。这是它值得单独写一篇文章的全部理由。

一次落在第④格的判定，形式审查全部通过：

- 诊断栏已填，无空缺；
- 病历完整率 100%；
- 诊断填写率 100%；
- 编码可入库，可统计，可结算；
- 若有质控，“病历完整性”与“诊断规范性”两项都满分；
- 若有一致性检查，多位医生对同一份资料会给出同一个证名——因为他们用的是同一个残缺的输出集，会被同样地填到同一处。

没有任何环节报错。而恰恰在这样一次判定里，那个位置本该是空的。

三条签名，逐条对照：

签名一：它在短期指标上表现为改善，而且是因果性的改善。这一条比“看起来好”更强。取消“不适用”这个选项，直接改善全部完备性指标。填写率、完整率、编码覆盖率、统计口径的完整性——每一项都因为“必须填”而变好。也就是说，指标的改善与填判的增加不是并存关系，是同一个动作的两面：要让填写率到 100%，就必须让每一栏都有可填的东西；而只要输出集里没有“不适用”，那 100% 里就必然含着填判。

签名二：它的失效不产生任何信号。这一条在本章所处理的形状里比在别处更彻底。填判的输出与正常判定的输出在形式上完全相同——不是“信号很弱”，是没有信号可言。同一个证名，同一栏，同一编码。任何下游程序、任何

统计、任何研究入组，都无法把两者分开。唯一会暴露它的场合，是这个判定在下游产生了明显的失败，而那时人们会去追问“是不是辨错了”——而那是一个错误的问题，因为它预设了有一个对的答案。

签名三：它自我加固，而且越正规化越严重。规范化的方向是明确的：结论要唯一、要可编码、要能进统计、要全覆盖、要可结算。每一项都在把“不适用”挤出输出集。一个尚未编码化的领域，医生还可以在病历上写一句“此人情情况特殊，暂难以某证概括”；一旦全面结构化，这句话就没有地方去了——**结构化的第一件事就是把自由文本换成下拉选项，而下拉选项里没有这一条。**

这里必须把话说清楚，否则本章会被读成反对规范化：**填写是必要的。**一个空着的诊断栏无法结算、无法统计、无法交接，而这三件事都关系到病人的实际利益。本章不主张允许空栏。本章主张的只有一件：

在输出集里加一项，并且给这一项配一条下游路径。

加一项而不配路径，比不加更坏——第十五节会说明为什么，那是本章被削掉的最重的一处。

九·弃判率：一个四领域量纲不同而比率相同的读数

一个判断若停在“有这么一样东西”，它还不能被检验。要能被检验，需要一个可以事后清点的数。

本章给出的读数是**弃判率**：

弃判率 = 一段时期内，判定输出中被记为“不适用／无从判定”、且带有下游处置路径的次数 ÷ 该期全部判定输出次数。

定义里的“且带有下游处置路径”必须写死，否则这个读数会被一个空字段架空。一个只能填、填了之后什么也不会发生的“无从判定”，不计入分子——它是第二种填判，只是换了个填法。**计入分子的条件是：这个输出被填之后，有一条规定的动作跟着它（转诊、扩展验证、另请判定、暂缓入组、单独列出）。**

它在四个领域里的清点方式，量纲各不相同，比率相同：

· **细胞学**：一张细胞图谱中，被标注为“未指认／中间态／不属已知类型”并保留在最终图谱里的细胞占比。**这里必须与质控剔除严格分开：被剔除的细胞**

不计入分子，因为剔除是删除而不是标注——它连同它提出的那个问题一起消失了。这个区分本身就是本章给细胞学的一条可操作建议。

- **工程学**：一份鉴定报告中，明确标注“本项工况未覆盖，使用前需另行验证”的条目数，占该报告所声明适用范围全部条目数的比例。
- **管理学**：审查与决策记录中，作出“本程序不适用于此情形，移交某处另行处理”结论的次数占全部审查次数的比例。
- **中医临床**：门诊与住院诊断中，记为“证候未明／本套判据不适用”且触发了规定动作的次数，占全部证候判定次数的比例。

四处的单位分别是“细胞”“工况条目”“审查次”“诊次”，量纲毫不相干；比率是同一个东西：**这套体系有多大比例的判定输出，是它承认自己不该判的。**

这个读数要立住，必须过四关：

第一关，无量纲。是比率，四个领域可以并排放在同一张图上比较。

第二关，可事后清点。不需要新仪器、不需要前瞻设计——它只需要**输出集里有那一项**。这是它与前两项读数最大的不同：留底率与存异率都可以对既有记录回溯计算，**弃判率不能**，因为在没有那一项的体系里，分子恒等于零。这不是这个读数的缺陷，**这正是它要指认的那件事**：一个恒等于零的读数，说明的是输出集缺一格，而不是不适用的情形不存在。因此它的正确用法是**先加格，再清点**，而“加格之后它是不是一直是零”本身就是最便宜的一次检验。

第三关，把一个印象变成一个数。“有些病人其实不该硬套证型”是一句人人有感觉、无人能拿它做事的印象。0.6%是一个数，可以排序、可以比较、可以进研究设计。

第四关，与既有主要指标正交。这一关最容易漏，也最容易毁掉一个新指标。

弃判率过这一关，而且是**反向过的**，而且反得极其干脆：

一个科室，病历完整率 100%，诊断填写率 100%，病案首页无空项，编码覆盖率 100%，证候诊断规范性质控连续三年满分——**它的弃判率必然是 0**。不是因为它做得差，是因为**这三年里它一次也没有被允许说“这个我判不了”**。

>

另一个科室，因为接收大量外院转来的疑难与多病共存病例，医生在病历上留下若干“目前难以某证概括，先按主症处理，另请会诊”的记录——它的**弃判率不为零**，而它的“诊断规范性”评分更低。

完备性指标越高，弃判率越低；而完备性指标是被硬性考核的。这不是相关，是**制度性的因果**：病案首页的诊断栏不允许空着，这是明文规定。**制度直接禁止了这个输出。**一个新读数能与现有指标形成这样清晰的反向关系，说明它不是任何现有指标的影子。

同时必须写死一条：**弃判率不是越高越好，而且它的滥用后果比另外两个读数严重得多。**一个高弃判率可能意味着这套体系诚实，也可能意味着推诿——把难的病人推出去，指标就好看了。**因此弃判率必须与转出后的去向一起读，单独看它没有意义。**第十七节把这一条写进了伦理红线。

十·一句不含情态词的问话

一个判断如果只能靠“应当”“充分”“真正”“恰当”来表述它的判据，它还没有落到可查的层面。本章的判据是一句可以直接拿去问表格与流程规范的问话，一个情态词也没有：

上一次复核的结论表上，“无从判定”有没有一个可填的格？过去一年被填过几次？

它问的是事实，不是判断。回答它不需要任何人的意见，只需要看表格和数记录。

把它在五个场合各跑一遍，答案互不相同——这一点很重要：如果五个场合答出同一个东西，那说明它只是承重命题的复述，不是一个能分辨情况的判据。

场合一，门诊病案首页。答案是：没有这个格，而且**规定不许空**。填写率是被考核的。这是最纯粹的第④格。

场合二，一张细胞图谱的注释表。答案通常是：有“unassigned”这一类，但它在发表的图谱里常常被合并掉或被隐去——因为一张有大片“未知”的图谱不好看，也不好讲故事。**格是有的，填了之后会被要求解释，于是倾向于不填。**这提示了一种此前没想到的中间状态：**格存在，而填它有社会成本。**这一条是查出来的，不是从命题推出来的。

场合三，一份型式试验报告。答案是：有——“未覆盖工况”是一个正式条目，而且写它对发证方有利（免责）。这是四家里唯一一处填这一格对填的人有好处的场合，因此它填得最实。这是一个很重要的发现：这一格能不能被填实，取决于填它的人得利还是受损。也是查出来的。

场合四，一次合规审查。答案是：多数没有；少数有“不予受理／移交”这一项，而它的使用被严格限制，因为它意味着审查未完成。格存在但被制度性地压制。

场合五，一次法庭立案。答案是：有，而且是成熟的——裁定不予受理、裁定驳回起诉、裁定移送管辖，每一种都有明确的法律后果、明确的文书格式、明确的下游路径，而且作出这样的裁定不被视为法官失职。这是四家之外的第五处，也是本章见过的最完整的第③格标本：一个“我这里判不了”的输出，配着一整套转出机制、书面理由与救济途径。

五个答案互不相同，其中至少三个（填它有社会成本、填它对填者有利、有格而被制度压制）不能从承重命题直接推出。而第三个答案给出了本章全部落地设计里最要紧的一条线索：让填这一格的人得利，而不是让他解释。

十一·复核路径的重排：第一问该问什么

现在可以正面回答那个问题了：一个证如何被确立并被复核，才算数？

现行的默认路径是两步：

确立 → 复核（对／不对）

本章主张它实际上是三步，而第一步现在没有位置：

判适用 → 确立 → 复核（对／不对／不适用）

三步各自的动作：

1. 判适用：在辨证之前先问一句——这套判据在这个人身上是不是一个可以问的问题。判据不是“我有没有把握”（那是不确定），是“他落在我的参照范围内吗”。可操作的问法在第十二节。
2. 确立：在适用的前提下辨证。这一步不变。
3. 复核：输出集从两项扩为三项。第三项必须配下游路径。

三个要点：

第一，可插手的那一步在最前面，而且它比另外两步都便宜。现行的全部改进努力压在第三步的准确性上——统一标准、加强培训、提高一致性、引入客观化检测。这些改进各自都对，而它们全部预设了第一步的答案是“适用”。在一个不适用的对象上把辨证做得更准，是把测量精度提高到一个错误的量上。这是本章与主流改进方向的分歧所在，必须说清楚：本章不反对提高辨证准确性，本章反对的是在没有第一步的情况下提高它。

第二，第一步的输出是三值而不是二值，这是全文的枢纽。“适用／不适用”如果只是二值，它会立刻退化为一新的判定，并且同样面临“这个判定不适用”的无穷回归。本章的处置是把它写成一个带阈值的连续量加一个转出动作：不是“这个人适不适用”，是“这个人离我的参照集中心有多远，远到什么程度就转出”。距离是连续的，动作是离散的，而阈值由科室自己定并公开。这样它不构成回归，因为阈值是一个约定，不是一个判定。

第三，第一步的成本几乎全部落在“转出去哪里”上，而不是“判不判得出”。判出一个人不在自己的参照范围内并不难——任何一位有经验的医生心里都清楚哪些病人“不好办”。难的是这句话说出来之后没有地方去。因此第一步的完整形态是：一个判据、一个阈值、一条转出路径。缺了第三样，前两样会在三个月内失效。

三门源学科的路径在这个框架下可以重新排列，各自的位置也清楚了：

- 细胞学走的是“从当下表达出发，经聚类，落到类型标签”；它的第一步被参数化掉了——分辨率与参照集的选择就是第一步，而它被当作一个技术参数而不是一次判定。
- 工程学走的是“从使用条件出发，经试验矩阵，落到适用范围”；它把第一步做得最好——它显式写下了作用域，而且写它对写的人有利。它缺的是使用端的消费点。
- 管理学走的是“从情形出发，经程序，落到符合与否”；它的第一步被制度性地压制——因为“不适用”意味着审查未完成，而未完成在管理上等同于失职。

三家里，工程学离本章的主张最近；而它离得近的原因很具体：它的作用域声明有免责价值。这是本章全部落地设计所依据的那条机制。

十二·七个可以立刻做的动作

一条判断如果给不出低成本的落地动作，它就只是一个说法。以下七条按成本从低到高排列。

一、在证候诊断的输出集里加一项：“本套判据在此例不适用／证候未明”，并明确它不计入“诊断不完整”类质控扣分。后半句比前半句重要得多——没有后半句，前半句在三个月内会变成一个从不被填的格。

二、给这一项配一条下游路径，至少三选一：另请判定、暂按主症处理并约定复评时点、转出并注明去向。弃判必须配转出，不得单独存在。一个没有下游的“不适用”，等于把病人扔在原地，比填判更坏。

三、在证候标准与量表的文本里，加一节“制定样本的构成与未覆盖范围”：地域、年龄段、病程阶段、合并用药背景、样本量。这一节在工程学的合格证上是强制的，在这里是空的，而加它的成本为零——这些数据在制定过程中本来就有。

四、让填这一格的人得利，而不是让他解释。这是第十节场合三给出的线索。可行做法：把“本例超出本科室常见谱系”作为一条**免责声明**而非一条**能力性声明**——它的效力应当类似于工程合格证上的适用范围，写了就减轻后续责任，不写则默认担保。这一条改变的是激励方向，是七条里最要紧的一条。

五、借法庭立案裁定的形式。不予受理、驳回、移送管辖，三种输出各有明确的文书格式、法定理由、下游路径与救济途径，而且作出这类裁定不被视为失职。这一整套形式对中医会诊与转诊完全适用，而且它已经运行了很久，证明了一个“我这里判不了”的输出可以被制度化而不损害体系的权威。

六、给“参照范围”设一个粗糙但可用的读数。不必精确：本科室近三年收治病例中，本例的主要特征组合出现过几次？为零或极低时提示可能超出参照范围。这个数从既有病案库里就能算，不需要新采集。**粗糙的判据配上明确的动作，胜过精确的判据配上没有动作。**

七、对已经结构化的电子病历，保留一个自由文本的“限定”栏，并在调阅时与诊断栏同屏显示。结构化的代价是把“此人情况特殊”这类话挤出了系统；留一栏并保证它被读到，是最低成本的补救。**注意：不与诊断栏同屏显示的自由文本栏等于不存在**——这是本系列问题另一篇文章已经指认过的失败方式。

七条里第一条与第四条最要紧：一条把格加上，一条让格被填实。**只做第一条不做第四条，得到的是一个从不被填的下拉选项。**

十三·与最近的几种说法的分界

一个新名字最常见的下场，是被读者正确地认出来是某个旧概念的改名。要避免这个下场，只有一个办法：把最容易与它混淆的说法逐个请进来，各自说清它到哪一步，再说清同一份材料上两边会给出什么不同的判断。

先把本章的命题剥掉中医与细胞学的一切名词，只留形式：

一个判定系统的输出集里没有“不适用”，因此所有超出适用范围的情形被填成了正常输出；填出来的结果与正常输出没有任何可区分的标记，而输出集的完备性正是被考核的东西。

这个纯形式的结构属于“越规范越失真”这一类反转结构，而这一类在管理学与组织社会学里早有名字：目标置换（Merton 1940）、成功陷阱与学习近视（Levinthal & March 1993）、核心刚性（Leonard-Barton 1992）。**必须先点名最贴的那一个。**最贴的是目标置换：手段（填写率）取代了目的（判断质量）。

分离线：目标置换的机制是**考核压力使人重视手段**——它要求有人在应付。本章的机制是**输出集缺一格**——即使没有任何考核、没有任何人想应付，只要输出集里没有“不适用”，填判照样发生。**判定形式：**在完全无考核、无绩效压力的场合——算法自动注释、无人监督的批量标注、纯科研目的的回顾性标记——若填判照样以同样的方向性发生，则本判断对而目标置换不足以解释；若在无考核场合填判显著减少，则应归于目标置换。**这一条可以直接做，而且算法自动注释正是一个零考核压力的现成场合。**

以下逐位。

（一）拒识选项与开放集识别（模式识别，Chow 1970；开放集识别，Scheirer 等 2013）

它到哪一步：这是本章最近、也最危险的方法学邻居。它明确指出分类器面对训练集之外的类别时会强行归类，并提出了给分类器加一个“拒绝”输出的整套方法：设置置信阈值、建模开放空间风险、评估拒识后的性能。**这一位几乎覆盖了本章的形式层。**

分离线：它是分类器层面的一个设计选项，本章是**制度层面**的账目位置。一个技术上具备拒识能力的分类器，若下游没有任何流程能消费“未知”这个输出——没有对应的处置路径、没有编码、没有付费方式、没有责任归属——那么拒识

输出会在使用中被人工兜底再填成一个具体标签，**填判照样发生，只是搬到了人工那一环。本章相对它的增量不在算法，在下游。**

判定形式：给一套判定系统加上拒识输出。若下游流程能消费它（有处置路径、有编码、有付费），则本章所指的空缺不存在；若下游仍强制要求一个具体标签，则拒识输出会被系统性地转回具体标签，本章对。**这一条可以在任何一家上线了辅助诊断系统的医院直接观察。**

（二）分布外检测与协变量偏移（机器学习，dataset shift 相关工作）

它到哪一步：它关心的是**检测**——能不能发现这个样本来自训练分布之外。方法很成熟。

分离线：它管发现，不管发现之后往哪里记。**一个分布外检测完美的系统，若没有“不适用”的账目位置，填判照样发生。**判定形式：在一个已部署分布外检测的场合，考察被标记为分布外的样本最终以什么形式进入记录——若仍以一个正常标签进入，本章对。

（三）外部效度与推广性（Campbell 与 Stanley 1963 一系）

它到哪一步：它问一个结论能推广到多远的人群、场景与时间，并提供了一整套威胁清单。

分离线：外部效度是一个**程度问题**，它的输出是“推广时需谨慎”；本章问的是**输出格式的有无**——在推广之外的那些个案上，系统给出的是什么。“需谨慎”不是一个可以填进诊断栏的东西。**判定形式：**一篇明确写了外部效度限制的研究，其结论在临床应用时，是否有任何机制在超出限制的病人身上产生一个不同形式的输出？若没有，本章所指的空缺独立于外部效度而存在。

（四）测量不变性与项目功能差异（心理测量）

它到哪一步：它检出同一个题目在不同群体上功能不同，处置是删题、分组计分或加参数。

分离线：它的所有处置**仍然输出一个分数**。没有“此人不可测”这个输出。**判定形式：**在一个已做过完整测量不变性检验的量表上，落在任何一组之外的个体（多重身份、边缘群体），最终得到的是一个分数还是一个“不可测”？若仍是分数，本章对。

（五）鉴定范围与型式试验（工程学）

它到哪一步：这是三家里唯一显式记录了作用域的一家，合格证上真的写着适用范围，而且**写它对发证方有利**。本章第十二节第四条的全部设计都借自这一位，必须明说这是借用而非发明。

分离线：它把作用域装在生产端，没有在使用端设检查——没有任何机制在超范围使用时报错，而超范围使用极其常见。本章相对它的增量只有一条：**使用端的消费点**。明写这一点，因为把一位近邻写成稻草人是最容易被拆穿的做法。**判定形式：**比较有使用端检查（如联锁、报警、准入校验）与只有证书声明两组，考察超范围使用的发生率。若两组无差异，本章的增量是多余的。

（六）“未特指”编码（疾病分类体系）

这一位必须重点处理，因为它是最容易被拿来说“我们已经有这个字段了”的一位。

它到哪一步：疾病分类体系里有“未特指”这一类，用于知道是这一大类、但无法细分的情形。它可填、可编码、可结算。

分离线：“未特指”是一个阳性归属——它断言“是这一类”，只是没细分。本章要的是**归属的悬置**——不断言是不是这一类。两者在账上完全不同：未特指的病例照样进入该类别的统计、照样可入组、照样按该类别付费；而一个真正的“不适用”应当**不进入该类别的分母**。

判定形式：查“未特指”编码的病例是否与明确编码的病例同样进入疗效统计与研究入组。若同样进入，则它不是一个拒识输出，是一个模糊的阳性归属，本章所指的空缺依然存在；若它被系统性地排除在该类别的分子分母之外，则它承担了本章所要求的功能。**这一条可以直接查数据字典，一天之内有结果。**

（七）“待查”与“疑诊”（临床习惯）

它到哪一步：临床有“发热待查”这样的成熟写法，它承认此刻判不了。

分离线：待查是**时间性的悬置**——现在不知道，检查推进就会知道，它有明确的闭合路径；填判所对应的是**结构性的不适用**——这套判据在这个对象上不成立，而时间与检查都不会改变这一点。**判定形式：**追踪一批“待查”记录与一批“证候未明”记录的闭合情况。若两者的闭合曲线形状相同（都随时间收敛），则本章所说的第三态不存在，它只是待查的一种；若“证候未明”这一类呈现出一个不随时间收敛的平台，则本章对。**这是一条可以直接跑的队列。**

（八）过度诊断（Welch 一系）

它到哪一步：发现了真实存在、但不会致害的病变，导致不必要的干预。

分离线：过度诊断是**判对了但不该判**（对象确实存在），**填判是判不了而判了**（对象是否属于该类不可知）。两者的干预方向相反：过度诊断要提高判定阈值，填判要增加输出选项。**对填判执行过度诊断的处方，只会让填判更严重——**提高阈值会把更多边缘对象推向“阴性”，而阴性同样是一个填出来的确定结论。

（九）有限理性与满意解（Simon 一系）

它到哪一步：它承认判断永远在信息不完全下作出，人只求足够好而非最优。这是管理学自己最接近本章的一位。

分离线：它说的是**判断的质量有上限**，本章说的是**判断的输出集有缺项**。二者可以同时为真而互不蕴含：一个完全理性、信息完备的判定者，面对一个落在参照集之外的对象，同样只能从现有输出里选一个。**判定形式：**在信息极其充分的场合（全套检查、多学科团队、充足时间），若边缘对象仍被填成一个具体证型，则本章对而有限理性不足以解释。

（十）分类的循环效应（科学哲学一系）

它到哪一步：分类会改变被分类者，因而分类需要修订。

分离线：它的机制需要被分类者**知晓**并作出反应，本章的机制不需要——一个细胞不知道自己被标成了什么，聚类算法照样把它填给最近的一团。**判定形式：**在完全无知晓、无反应的对象上（细胞、器件），填判是否照样发生？答案是肯定的，这也正是本章选择细胞学作为一家的理由之一。

十三之一·划完之后：这些邻居各自看不见什么

把上面每一位再问一遍“它把什么当作给定，因此结构性地看不见什么”，句式固定：

- **开放集识别**把“总有一个真实类别存在，只是我们可能认不出”当给定，因此看不见“这个对象不属于任何类别”的情形——一旦承认后者，它的评估指标就没有了参照答案。
- **疾病分类体系**把“每一次就诊都必须产生一个可编码的结论”当给定，因此看不见不可编码的就诊——一旦承认，结算与统计的口径都要重建。
- **型式试验**把“作用域由供方声明即可”当给定，因此看不见使用端的超范围使用——一旦承认使用端必须校验，责任划分要重写。

- **合规审查**把“程序总是适用的”当给定，因此看不见程序不适用的情形——一旦承认，审查就可能无法完成，而无法完成在管理上等同于失职。
- **细胞图谱**把“每个细胞属于某一型”当给定，因此看不见连续谱上不属于任何型的细胞——一旦承认，“细胞类型”这个概念本身要重做。
- **外部效度**把“结论总是可以被推广，只是范围有大小”当给定，因此看不见“在此处根本不构成一个问题”的情形。

六行背后是不是同一块地？是。它可以写成一句：

每一个被提出的问题都必须产生出一个答案；“这里问不出来”不算一种回答。

本章把它叫做**必答假定**。没有任何一派会去质疑它，因为互相批判的各派都站在它上面：主张更严格量化的与主张专业判断的，主张全覆盖编码的与批判编码化的，全部预设了一次判定必须产生出一个可用的结论。分歧只在“用什么方法产出”。“未被任何学科辨识”的真正来源不是没有人看，是看的人都站在它上面。

本章所说的那一份，正是从这块地的缝里掉下去的：它既不是正确的判定，也不是错误的判定，它是在**必答假定**下被生成出来的那个形式合格的答案——而“被迫作答”在所有账本上都不是一个会产生条目的事件。

十四·与本系列各章的分界

本系列近期有数篇处理相邻问题，其中两篇与本章同题，三篇与本章同学科组。它们是本章最近的邻居，必须逐篇划清。

《并身》（第四章）（教育学×化学×艺术学，与本章同题）。那一篇说的是：上一版判断通过它所授权的执行改写了复核所要面对的对象，而这份改变被读作对象自身的属性；读数是留底率。

分界：那一篇的缺栏在**对象侧**（被判断者已被改写），本章的缺栏在**输出侧**（判定体系缺一格）。二者完全独立，可以取到四种组合。**判决性对照**：一位从未接受过任何治疗的初诊病人，留底率为1、并身为零——若他落在这位医生的**参照谱系之外**，**填判**照样发生。反过来，一位被反复治疗改写、留底率接近零的病人，若他恰是这位医生最熟悉的那一类，则**填判**不发生。两个机制的启动条件毫不相干。

《摊差》（第五章）（物理学×心理学×工程学，与本章同题）。那一篇说的是：复核必然产出一份任何一方都消解不了的分歧，而这份分歧被吸收进“不确定度／共识／已接受”三只筐之一；读数是存异率。

这一篇与本章形状最近，必须划得最细。两篇都在说“某个状态没有账目位置”，而触发条件完全不同：

- 摊差的前提是有多个判断且互相矛盾。没有分歧，就没有摊差。
- 填判在只有一个判断、无任何分歧时照样发生。一位医生独自看诊，无人反对，他给出的证名照样可能是填判。

判决性对照，这一条是两篇之间最硬的一刀：在一次存异率为1的会诊中——三位医生的分歧被完整留痕、技术内容俱全——若这三位医生一致地把一位落在他们共同参照范围之外的病人判成同一个证，则摊差为零而填判为真。一致性越高，摊差越少而填判越隐蔽。两个机制在一致性这个变量上方向相反，这使它们可以被同一份材料分开。

叠加效应：两者同时发生时，一次复核既吃掉了分歧又填了不该填的格，而账面上会呈现为一次高度一致、完整、规范的判定。第八节签名一在这里达到极致：三个读数（留底率、存异率、弃判率）全部为零的一次复核，在现行的一切质控标准下是满分。

《叠裕》（《治未病立账学》第三章）（同三学科，处理“何谓未病”）。那一篇说的是：各单位全合格，而这些合格依据同一份不可分割储备的重复整份计入；读数是叠计比。分界：那一篇处理余量按单位分账，本章处理判定的适用范围。判决性对照：一个叠计比为1（各单位余量完全独立、无任何共用）的系统，其判定照样可能全部落在参照集之外——独立性与适用性是两件事。

《漏汰》（《治未病立账学》第六章）（同三学科，处理“未病如何被识别、无症状时如何正当干预”）。那一篇处理的是该被淘汰的没有被淘汰——识别侧漏掉了一次否定。本章处理的是该说不适用的说了适用——复核侧多做了一次肯定。两者方向相反：一个是缺了一次否，一个是多了一次是。判决性对照：在一个淘汰机制完全有效、无任何漏汰的系统里，填判照样发生——因为淘汰机制处理的是“该不该留下”，不处理“这套判据该不该被应用”。

《功债》（《治未病立账学》第九章）（同三学科，处理“治未病为何实践最少”）。那一篇说的是：预防不消除扰动，而是把它转成必须一直被供养的残留；读数是续养率。分界：功债是做了之后放不下来，填判是本不该做而做了。

二者在时间上一前一后：填判发生在动作之前的判定环节，功债发生在动作之后的维持环节。**判决性对照**：一次基于填判而启动的干预，其功债与一次基于正确判定而启动的干预**在量上没有系统差异**——功债只看动作，不看判定对不对。这说明两个机制作用在不同的环节上。

把六篇并排看，它们共有一个形式——**发生了而没有产生条目的东西**。这个形式的生产力太高，本身就是警号。本章与《并身》《摊差》同题，三篇一起排在被重估名单的最前面。而本章比另外两篇更没有资格例外：**它自己刚刚论证过，一个必答的体系会在任何位置上产出形式合格的答案——而“用这个形式去看任何领域都能看出点什么”，正是必答假定在方法论层面的表现**。第十八节要回到这一点。

十五·三门学科反过来给本章加的约束

借用不是单向的。三门学科在被本章使用的同时，也各自砍掉了本章原本打算主张的一些东西。三处削减都写在这里，并写清削掉的是什么。

第一处，管理学削掉了本章最自然的那个主张。

本章原本会写下这样一句更强的话：“**应当允许判定者说‘我判不了’。**”这句话听上去只是一个诚实的要求，成本很低，而且道德上无可指摘。

管理学不允许它成立，而且理由很硬：**判断必须终止并给出结论，因为要有人负责**。一个“不适用”的输出，在管理上等于把责任推回给提问者。而在大量场合这是不可接受的——急诊、灾难响应、必须当场决断的时刻，一个“我判不了”不是诚实，是把人扔在原地。管理学几十年处理这个问题的方式很清楚：**允许判断有限，但不允许判断缺席；缺席必须有接替，接替必须有指定**。

因此本章的主张必须削弱为：**弃判不是善，弃判加转出路径才是善；一个没有下游的“不适用”输出，比填判更坏**。这是一处实质削弱：它把“加一个选项”这个近乎无成本的要求，换成了一个必须配套转诊机制、责任交接与救济途径的制度设计，成本大得多。**而且它使本章的主张在缺乏转出资源的基层场合不可实施**——那里恰恰是填判最多的地方。这一处削减很痛，因为它削掉的正是本章最想说的每一句。

第二处，工程学动了本章的落地设计与激励方向。

本章原本会写下这样一句更强的话：“给每个判定标注适用范围，问题就解决了。”这句话顺着第二节工程学那一段很容易滑出去。

工程学不允许它成立，因为它做了，而且做了几十年。它的实际经验是：适用范围的声明会被写得尽可能宽泛——写窄了限制销售、限制适用场合、招来质疑；而使用端几乎没有人去读它。一份合格证上的温度范围，在采购、安装、运行的任何一个环节都很少被真正校验。标注在生产端是廉价的，消费在使用端是昂贵的，而制度的注意力全在廉价的那一端。

因此本章必须削弱为：标注作用域是必要而远不充分的；决定成败的是使用端有没有一个消费点，以及标注它的人得利还是受损。这一处削减改变的是本章的落地重心——从“加一栏”移到“改激励”，而后者难得多。第十二节第四条整条都是这一处削减的产物。

第三处，细胞学把本章从二值改成了连续，动的是命题的形式本身。

本章原本会写下这样一句更强的话：“每个判据都有一个明确的适用范围，对象要么在里面要么在外面。”

细胞学不允许它成立。它自己的证据说明：细胞类型之间不是离散的岛屿，是连续谱上被人为切开的区段；中间态是普遍的，不是例外；分辨率参数一动，边界就动。根本不存在一个客观的“里面／外面”分界。

因此本章必须放弃“适用范围有明确边界”的表述，改为：适用度是连续的，而输出格式是离散的——问题出在这个不匹配上，而不是出在边界画错了地方。这一处改动看似技术性，实则很重：它意味着本章不能主张任何“客观的适用性判据”，只能主张“一个由使用者公开设定并承担后果的阈值”。也就是说，本章所要求的那一格，其填与不填本身也是一次判断，而不是一次测量。承认这一点，比假装能给出客观边界诚实。

十六·不适用的边界

本章的命题在下列情形不适用，援引者应当自己先排除：

一、判定对象高度同质、参照集稳定的场合。多数常规、成熟、低歧义的判定落在第②格，从未出事，也不需要本章。为这类场合加一格是纯粹的成本。

二、必须当场决断且无转出可能的场合。急诊、灾难现场、资源极度受限的基层。此时填判是正确的临床选择——在没有下游的地方引入弃判是有害的。本

章对这类场合能提供的只有一条：**事后补记**，而补记会损失内容，这是必须承认的代价。

三、判定不指导任何动作的场合。 若一个判定只是描述性的、不据以行动，填判的后果有限。

四、本章交出去的第一个洞：那一格的填与不填本身也是一次判断，因而面临同样的问题。 一位医生判断“这个人在我的参照范围之外”，这个判断本身不适用？本章的处置是把它做成一个公开的阈值约定而非一次判定（第十一节第二点），但这只是把问题从“判断”移到“约定”，并没有消除回归，只是让回归停在一个可以被公开讨论的地方。这不是一个解决，是一个安排。

五、本章交出去的第二个洞：弃判率在没有那一格的体系里恒等于零，因而它无法用于回溯研究。 这与本系列另外两篇给出的读数不同——留底率与存异率都可以对既有记录回溯计算。**弃判率必须先改制度再测量**，这使它的检验成本显著更高，也使它更容易被“先做个试点看看”这类安排无限期推迟。

六、本章交出去的第三个洞，也是最大的一个：本章说不出填判的量级。 全文论证了填判的存在与机制，却给不出任何关于“临床上有多少比例的证候判定是填判”的估计——因为估计它需要一个本章所要求的、目前尚不存在的输出格式。一个说不出量级的机制，无法与“这只是极少数疑难病例的问题”这种反驳区分开。本章对此没有办法，只能把它写在这里。

十七·证伪条件、样本纪律、伦理边界，与一条写死日期的赌注

十七之一·最强的竞争解释

在给出证伪条件之前，先把那三句最难反驳的话摆在明处：

“这其实不就是填写率被考核、大家在应付吗？” “这其实不就是医生水平不够、见得少吗？” “这其实不就是罕见病例本来就难判吗？”

一条证伪条件如果不能排除这三句，它只是在验证“本章描述的现象存在”，那不算数。

条件一（核心，机制证伪）：控制判定者的资历、病例的客观复杂度、可用检查项数之后，若“输出集是否含‘不适用’一项”与“边缘对象被判入密集证

型的比例”之间不存在剂量-反应关系（含中间形态：有该项但计入质控扣分／有该项且不扣分但无下游路径／有该项且有下游路径），则本机制为伪——届时填判应归因于判定者能力或病例难度。

条件二（排除考核压力，即目标置换）：在完全无考核、无绩效压力的场合——算法自动注释、纯科研目的的回顾性标注、无人监督的批量标记——若填判照样以同样的方向性发生（系统性偏向参照集中的密集类别），则本判断对；若无考核场合填判显著减少，则应归于目标置换。**这一条可以立刻做，且成本极低。**

条件三（排除能力与难度）：在信息极其充分的场合（全套检查、多学科团队、时间充裕、判定者为高资历），若落在参照集之外的对象仍被系统性地填成具体证型，则不是能力问题。**具体设计：**取同一批边缘病例，分别交给两组高资历医生，一组的表格含“不适用”项且不扣分，另一组不含。若两组的证型分布无差异，本章为伪。

条件四（区分待查，排除时间性悬置）：追踪一批“待查”记录与一批“证候未明”记录的闭合曲线。若两者形状相同（都随时间与检查推进而收敛），则本章所说的第三态不存在，它只是待查的一个变体；若“证候未明”呈现一个不随时间收敛的平台，则本章对。**这是一条可以直接跑的队列研究。**

条件五（正向读数，顺序预测）：在引入“不适用”选项+不扣分+下游路径的机构，该选项的实际使用率上升，应当出现在“填它对填的人有利”这一条被明确之后1-两个周期，而不是与选项上线同步。若使用率在选项上线当月即上升并保持，则改善应归于宣贯的注意力效应；若上线后长期为零、直到激励条款明确才上升，则本章第十二节第四条的机制得到支持。**顺序比相关难被巧合满足得多。**

条件六（分层引用与分层证伪）：本章的主张分三层。第一层是“填判”这个构念（最强、最易倒）；第二层是那张两轴四格的辨别表（作为分析工具，即使第一层被推翻，第③格与第④格的区分仍可用）；第三层是一条不依赖前两层的经验主张——“在同一份边缘病例材料上，表格含‘不适用’项的判定者所给出的证型分布，与表格不含该项者显著不同”。第三层最稳、最便宜、也最容易检验；若第三层被推翻（两组分布无差异），本章全盘作废。

十七之二·样本纪律

上述条件涉及跨案例比较的，样本不得少于六个同质单元，且在关键混淆变量上同质。可用单元：同一医院不同科室、同一学会不同工作组、同一平台的多个标注团队、同一数据集的多次独立注释。严禁拿两家医院或两个地区的对比充数——两边在病种构成、信息化程度、质控文化上处处不同，所谓“控制掉某变量”在方法上站不住；这类比较只能作启发式例证，不能充当证伪检验。

十七之三·伦理与证据边界

本章的读数比本系列另外两篇更容易被滥用，因为它直接触及“拒绝判断”这个动作。以下四条必须写死，缺一条本章不宜被引用：

一、**弃判必须配转出，不得单独存在。**这是本章全部伦理条款里的第一条，也是唯一一条红线。一个没有下游的“不适用”输出，等于把病人扔在原地。任何援引本章而只加选项、不建路径的做法，都是对本章的误用。

二、**不得为提高弃判率而拒绝可以判断的病人。**弃判率一旦进入考核，最直接的滥用就是推诿：把难的、慢的、付费低的病人判为“不适用”转出去，指标好看而病人受损。因此弃判率必须与转出后的去向、以及转出病人的最终归宿一起读，单独看它没有意义。

三、**这个读数指的是位置，不是人。**弃判率低不说明这位医生保守或不诚实，它说明的是这套体系没给他这个输出。尤其不得用于评价个人。

四、“不适用”的记录不得用于事后追责。若这一栏可以被用来证明“你当时就知道自己判不了却还是治了”，它会在三个月内变成空栏。制度上必须明确该栏的用途限于处置与学习。

十七之四·反噬预言

本命题一旦被制度采用，最省事的达标方式不是弃判，而是把“证候未明”当成一个新的证型来用。一旦它可填、可编码、可结算、可入组，它会迅速膨胀成一只新的筐：难判的往里扔，不想判的往里扔，赶时间的往里扔。而那时它自己就成了一个填判——一个形式上合格、无法与真正的不适用区分的输出。

这个反噬有一个自指的形状，使它格外难解：本章所要求的那一格，一旦被制度化，就获得了它自己的密集区引力，于是它开始吸收本不属于它的对象。本章全文论证的机制，会原样作用在本章的处方上。

我看不到出口。唯一能想到的部分缓解是把这一格设计成**不可结算、不可入组、且必须配转出记录**——让填它没有收益只有免责，从而只有真正需要它的人才填。但这又与第十二节第四条“让填的人得利”相冲突：免责是一种利，而任何一种利都会诱发过度使用。**两边不能同时优化**。把这一条写在这里，是因为一个看得到出口的反噬不是反噬，只是待办事项。

十七之五·一条写死日期的赌注

到二〇三四年十二月三十一日，至少有一部由国家级中医药行业组织颁布的病历书写规范、诊疗规范或证候标准中，把“本套证候判据在此例不适用／证候未明”列为一个正式的、可填的结论选项，并同时规定：该选项不计入诊断完整性类质控扣分，且配有明确的下游处置路径。

什么不算命中，五条写死：

1. 仅出现在教材、指南讨论稿、专家建议或学术论文中，未进入颁布文本的，不算；
2. 仅允许写“待辨”“待查”“证候不典型”而无下游处置路径规定的，不算——这一条尤其要紧，因为这类写法在现行病历里已经存在，而它正是本章所指的时间性悬置，不是结构性不适用；
3. 虽设该选项但仍计入诊断完整性扣分的，不算——在这种安排下它是一个永远为零的格；
4. 该选项被赋予独立的结算编码或研究入组资格的，不算——那说明它已经变成了一个新的证型，即本章所预言的那个反噬；
5. 由本章作者或其所属机构参与起草的文本，不算。

十八·本章自己是不是它所描述的那样东西的标本

按本章的判据，本章自己落在哪一格？

先问一句：本章这套判据，在它自己身上适用吗？

本章的判据是：一个判定体系的输出集里有没有“不适用”这一项。本章是一篇学术论文，而学术论文这个体裁的输出集里没有这一项。一篇论文必须提出一条承重命题，必须论证它，必须给出结论。没有一种论文格式允许作者写“这个问题在我这里问不出来”并就此交稿。期刊不收，栏目不排，读者不读。

于是本章正落在自己所指认的第④格里，而且落得比本系列另外两篇更深：另外两篇指认的是记账缺栏，本章指认的是**输出集缺项**，而本章自己的输出集缺的正是同一项。

这三条具体后果：

第一，本章对三门学科的处理，每一处都可能是填判。例如第二节把管理学的立场概括为“算数落在程序上”——管理学内部对此有大量争论，而本章必须给出一个可用的概括，因为“**管理学在这个问题上的立场无从概括**”不是一个可以写进第二节的输出。同样，第六节把“结合临床，随证治之”读成一种填法——这句话确实可以这样读，但它是否本该这样读，本章自己分不出来，而本章的格式要求它给出一个读法。

第二，本章最不该被引用的方式，是被当作已经证成的东西引用。本章能提供的只有第十七节条件二与条件三那两条——一个成本极低、任何人可以独立做出来的对照。在那个对照被独立完成之前，本章的地位是一个待检验的读数，不是一个结论。

第三，本章的弃判率是零，而且这个零不是装出来的谦虚，是可以被清点的。全文十八节，没有任何一处写下“此处本章无从判定”。第十六节列了三个洞，但那是已知的局限，不是输出的悬置——“本章说不出填判的量级”仍然是一个明确的判断，不是一次弃判。要在这篇文章里真正弃判一次，需要在某一节写下“这一节所问的问题，在本章的框架里不成立，本章不给结论”，而本章没有做到。做不到的原因，正是本章第八节签名三所描述的：格式要求完备，完备要求填满。

对此本章只能做三件事，而且都已经做了：第十五节写清三处削减、第十六节交出三个洞、第十七节把最便宜的两条对照单独列出并写明它们不需要本章作者参与。一篇落在第④格里的文章，唯一的自救方式是把能推翻自己的那件事做得足够便宜。

最后一句，关于这一整批文章。本系列近年数篇——包括本章与它的两篇同题姊妹篇——共有一个形式：找出一样“发生了而没有产生条目的东西”。这个形式的生产力太高，本身就是警号。而本章比其他几篇更没有资格例外，因为它自己刚刚论证过一件事：一个必答的体系，会在任何位置上产出形式合格的答案。一把在任何门上都拧得动的钥匙，通常不是钥匙，是撬棍；而一套在任何领域都能“看出缺栏”的读法，很可能正是这套读法自己的填判。若有人能证明这

个形式在任意领域都能开出一样看似深刻的东西，那么这一批文章都要一起重估，本章与它的两篇同题姊妹篇排在最前面三位。

十九·伦理与证据边界声明

本章提出的弃判率是一个结构读数，不是绩效指标。**弃判必须配转出，不得单独存在**——这是本章的第一条红线，任何只加选项、不建路径的援引都是误用。禁止用于个人考核与晋升评价；禁止为提高该读数而拒绝可以判断的病人；“不适用”的记录不得用于事后追责，其用途应限于处置与学习；据该读数作出的任何不利安排，须公开编码规则并提供复核通道。本章所举临床情形均为类型化描述，不涉及任何具体病例、机构或个人。本章不构成任何诊疗建议，也不构成对任何已颁布标准或分类体系的技术评价。

注释

1. 本章所称「证」指中医辨证论治体系中的证候判断；所称“复核”指对已确立之证的再次判断，含同一诊次内多人评判与不同诊次的前后评判两种情形。
2. 第二节对三门学科验证范式的描述取其主流形态，不代表该学科内部无争议；三家之间的冲突为本章所整理，各家并未如此互相表述。
3. 第四节对细胞类型注释流程的描述取其操作结构，具体算法、参数选择与校正方法详见相关方法学文献；本章不评价任何具体做法在其领域内部的适当性。
4. 第九节弃判率的定义中，“带有下游处置路径”是计入分子的必要条件；缺此条件的记录不计入，理由见第十二节第二条。
5. 第十节场合五所述司法文书形式，本章只取其形式对照，不涉及任何司法制度评价。
6. 第十七节赌注的判定以颁布文本的正式发布日为准；五条排除项须同时满足方可判为命中。

第三编 还原

为什么七十年还原不成而临床有效

第七章 随程

——为什么每一轮还原研究结束的时候，它要还原的那个对象已经被这一轮研究自己重划过一次

提 要 「证」自二十世纪五十年代进入实验室以来，历经内分泌、免疫、代谢、基因、蛋白、代谢组与人工智能各代技术，始终没有被还原为一组稳定的、跨研究可重复的物质指标；与此同时，以辨证为前提的处置在临床上持续有效，这一有效性并未随还原的失败而衰减。本章主张：不可还原与持续有效不是需要分别解释的两件事，而是同一个机制的两侧。「证」不是一个尚未被测准的隐藏实体，不是一套评分者信度不足的主观编码，也不是一种原则上不可言说的整体把握，而是一类只在某一套识别程序运行期间才具备同一性、并且会反过来改写该程序的对象——本章称之为**随程物**。它有效，是因为它的划分与当下可执行的处置手段清单共用同一套切法；它不可还原，是因为每一轮还原研究在结束之前，都会通过修订识别程序而把它自己的对象重划一次。本章取分析化学的操作定义形态分析、教育诊断学的可教学性判据、文物与艺术品保存的作品同一性三条互不相容的驱动主张相互顶撞，用分析化学自己在顺序提取标准化中已经掌握、却从未往这个方向使用的一件事实——被认证的程序反过来规定了下一批标准物质与下一轮提取步骤——推翻三家共同假定的「驱动方向固定」。本章给出一个跨领域量纲相同的读数**随程率**：最近 N 次对象归属变更中，由程序或标准变更引起、而非由对象出现新读数引起的次数占比；并给出一句不含情态词的判据：**最近 N 次分类变更中，有几次当日的记录里没有出现新的观察项？**全文的可裁决主张是：证候分类的重大重划在时间上紧跟处置手段库的变动，而不紧跟病理机制与检测技术的进展；若在六次以上的标准修订中这一顺序关系不成立，本判断为伪。

一·引言：一件持续了七十年的怪事

一九五九年前后，上海的一间实验室开始测量一批被中医判定为「肾阳虚」的病的人的二十四小时尿十七羟皮质类固醇。读数偏低，且与对照组分得开。这是「证」第一次被交到一台仪器手里，也是七十年工作的开端。此后每一代新技术出现，都有人立刻把它对准同一个对象：内分泌之后是免疫，免疫之后是自由基与微循环，然后是基因芯片、蛋白质组、代谢组、肠道菌群，最近十年是机器学习与多组学整合。每一代都产出过分得开的读数，每一代都有过「找到了物质基础」的宣告。

七十年之后，情况是这样的：任取两项关于同一个证型的标志物研究，它们给出的指标集合重叠得很少；把时间拉长，重叠率也没有随着样本量和技术世代的提升而单调上升。这不是某一项研究做坏了，而是一个稳定的整体形势。医学史上并非没有难缠的对象，但难缠的对象通常呈现另一种形势：早期众说纷纭，随着方法改进而逐渐收敛，最后落在一两个指标上。幽门螺杆菌走的是这条路，糖化血红蛋白走的是这条路，肺栓塞的 D-二聚体也走的是这条路。「证」不走这条路。它的表现不是收敛得慢，是**不收敛**。

与此同时，还有另一半事实。以辨证为前提的处置，在临床上并没有因为还原的失败而失效。一个中医师问诊、切脉、看舌，判定一个证，然后开方；病人好转。这个过程在数以亿计的诊次里被重复。它当然含有自愈、安慰、回归均值与选择偏差，但把这些全部扣除之后，仍有一部分剩余是稳定的——否则不会有任何一个国家的医疗体系愿意在其中长期投入。更要紧的是一个细节：**同一个病人，换一位医师，可能被判为另一个证，开出另一张方，也可能好转**。这个细节在七十年的还原研究里几乎从未被当作正面的证据使用，它一直被当作噪声、当作信度问题、当作有待规范的乱象。

于是问题成形了：**一个七十年测不准的东西，凭什么一直管用？**通常的处理是把它拆成两个问题分头解决——先解释为什么测不准（技术不够、对象复杂、标准不统一），再解释为什么管用（部分成分有药理活性、心理效应、经验规律）。本章主张这个拆法本身是错的。**不可还原与持续有效不是两件事，是同一个机制的两侧**；只要把它们分开处理，就永远解释不了为什么两者能同时维持七十年而互不侵蚀。

为什么现在问这个问题？因为过去二十年出现了两样以前没有的东西。其一，证候诊断标准的修订已经积累到足够多的次数，可以被当作一个时间序列来看，而不再只是背景。其二，别的学科在同一时期各自遇到了结构相同的麻烦，并且各自给出了强硬的、互不相容的解决方案——分析化学处理「一份土样里的重金属究竟以什么形态存在」，教育诊断学处理「一个学生究竟属于哪一类」，文物与艺术品保存处理「这件东西究竟还是不是那件作品」。这三门学问离中医很远，远到不可能互相引用；但它们各自都被逼到了同一个墙角：**一个对象的身份，究竟由什么决定**。

本章要立的判断是：存在一类对象，它的同一性只在某一套识别程序运行期间成立，并且它会反过来改写这套程序；「证」是这类对象里被观察得最久的一

个样本。本章称这类对象为**随程物**。走法是：先把三种现成解释摆出来，指出它们共同站着的那块地；再从三门外学科各取一条互不相容的驱动主张，让它们互相顶住；然后用其中一门自己已经掌握、却从未往这个方向使用的一件事情，把三家共同的假定推翻；由此得到承重判断、两根轴、一个可清点的读数、一句不含情态词的判据；最后交出适用边界、证伪条件、反噬预言和一条写死日期的赌注。

二·把问题重新问一遍

先把「还原」这个词说清楚，否则后面全是空转。

七十年里的还原工作，不管用什么技术，形状都是同一个：**先由中医专家按某一版标准把病人判定为某证，形成一个组；再在这个组与对照组之间寻找差异指标；找到的差异指标被称为该证的物质基础。**这个形状有一个前提是从不被写出来的——**分组是给定的，指标是待找的。**也就是说，还原研究把「证」当作自变量，把指标当作因变量；「证」这一侧被默认为稳定的、外生的、由病人身上正在发生的事情单向决定的。

现在把七十年当作一整段时间来看，这个前提就出问题了。因为在这七十年里，「证」这一侧一直在动，而且是被研究者自己动的。一九八六年有一版全国性的虚证辨证参考标准，一九九三年有一版新药临床研究的证候指导原则，一九九七年有一版国家标准的证候术语，二〇〇二年指导原则修订，其后有证素辨证一路的重新拆分，二〇二一年国家标准再修订，其间还有各病种的行业标准、专家共识、临床路径若干。每一次修订，都改动了「哪些表现算数、算到什么程度、如何组合成一个证名」。

于是七十年还原工作实际上是这样跑的：第一轮研究用一九八六年的分法分组，找出一组指标；第二轮研究用一九九三年的分法分组，找出另一组指标；两组指标重叠很少。通常的解释是「方法学质量参差」。但还有一个解释更简单：**两轮分的不是同一个东西。**

这个解释一旦说出口，问题的形状就变了。原来的问题是「这个对象为什么这么难测」，新的问题是「**这个对象在两次测量之间被谁改了**」。前者是一个精度问题，后者是一个本体问题。

同样的重问也适用于有效性那一侧。原来的问法是「辨证论治的疗效来自哪个成分」，答案候选是药理活性、心理效应、经验规律。新的问法是：**在一个对**

象会被识别程序改动的情形下，「有效」是相对于哪一次划分而言的？如果一位医师按二〇二一年的标准把病人判为甲证，开甲方，病人好转；另一位医师按一九八六年的标准把同一位病人判为乙证，开乙方，病人也好转——那么有效性显然不是「命中了病人体内那个固定的东西」。它更像是另一种东西：**两位医师各自用一套完整的、自洽的划分把病人切了一刀，而每一套划分里都配着与之匹配的处置手段。**

这就把两侧接上了。若对象随程序重划，那么每一套程序内部都可以是有效的，而跨程序的还原永远不收敛。这句话现在只是一个可能性，它需要一个机制来支撑，需要说清楚是什么在驱动这种重划、谁先动、下一轮换不换人。这正是本章后面要做的事。

在往下走之前，先把三种最常见的现成解释摆出来，因为本章的判断如果不能把它们各自的位置说清楚，它就只是第四种说法。

三·三种现成解释，以及它们共同站着的那块地

三之一·隐藏实体说：还没测准

第一种解释最朴素，也最有力量：「证」对应着病人体内某种真实存在的状态，只是这种状态涉及的层次太多、变量太密、个体差异太大，现有技术还够不着。等到多组学数据足够、算力足够、队列足够大，它会被刻画出来。

这种解释的合理性来自医学史本身：太多曾经被认为不可测的东西后来都被测出来了。它给出的处方也很清楚——加大投入、扩大样本、提高维度。七十年里绝大部分资源投在这条路上。

它的困难在于，它对「不收敛」这个事实没有说法。若只是技术不够，那么随着技术进步，跨研究的指标重叠率应当上升；一个持续七十年、跨越六七代技术而重叠率不升的形势，用「还没到时候」是解释不了的。它只能一次次把兑现日期往后推。

三之二·主观编码说：信度不够

第二种解释把矛头指向识别端：「证」是医师的主观判断，不同医师之间一致性有限，所以以此分组的研究必然噪声巨大。处方也很清楚——规范化。统一

术语、统一权重、统一操作定义、做量表、做四诊仪、做人工智能辅助辨证，把评分者一致性拉上去。

这条路在过去三十年做出了实打实的成绩。若干病种的证候诊断标准把一致性系数从零点六左右拉到零点八以上，这是可以查证的进步。

它的困难在于两点。第一，一致性上去了，跨研究的指标重叠率并没有跟着上去——如果噪声是主因，减噪应当带来收敛。第二，也是更麻烦的一点：**每一次成功的规范化，都伴随着一次大规模的重新归类。**同一批存档病例，按新标准回溯判定，相当比例会换一个证名。若「证」是一个客观状态而标准只是量尺，那么换量尺不应当换类；换类说明被量的东西跟着量尺变了。主观编码说没有位置安放这件事，它只能把它记为「新旧标准的差异」，作为方法学细节写在讨论部分。

三之三·默会知识说：本属另一种知识

第三种解释来自另一个方向：「证」根本不是命题性知识，而是一种在实践中生成、在师徒之间传递、原则上无法完全形式化的把握。名医的手感、脉下的分寸、看一眼就知道的那种判断，都不是可以拆成条目的东西。按这个读法，七十年的还原之所以失败，是因为一开始就找错了知识的类型；而它之所以有效，正因为它保留了形式化会破坏的那一部分。

这条解释在人文一侧有深厚的支持，从波兰尼关于默会之知的论述，到人类学对中医临床实践的长期观察，都指向同一个方向：知识长在实践中，教科书化会让它流失。

它的困难同样是一个可以查的事实：**规范化并没有带来疗效的整体下降。**按新标准培训出来的一代医师，用条目化的辨证流程处置，其疗效并没有塌方式地低于前一代。如果真正起作用的是不可言说的那一份，而规范化正在挤掉它，那么疗效应当随规范化程度上升而下降。这条预测没有兑现。默会知识说必须把它解释成「流失还没来得及显现」或「新的默会知识又长出来了」——后一种说法其实已经很接近本章的判断，但它没有把这一步走完，因为它把重点放在**知识的存在方式**上，没有去问**对象的存在方式**。

三之四·三家共同站着的那块地

把三种解释并排放，它们分歧极大：一家把问题归给仪器，一家归给识别者，一家归给知识的类型。处方也相反：加大投入、加强规范、保护师承。但三家共有一条从不被说出的假定：

在任何一轮里，「证」这一侧与「识别程序」这一侧之间，谁决定谁是给定的。

隐藏实体说假定病人身上的状态单向决定证，识别只是读取；主观编码说假定客观状态单向决定证，识别的误差是要被压掉的干扰；默会知识说假定实践中的整体状况单向决定证，形式化只会损耗。三家吵的是**读取端有多可靠、损耗有多大、要不要相信它**，从没有一家怀疑过**读取的方向**。

把这条假定念给三家听，三家都会说：这还用说吗？一个病人身上正在发生什么，当然不取决于我们拿什么册子去看他。

这句「这还用说吗」，正是本章要动的那一处。要动它，需要三样东西：三条互不相容的、各自主张「驱动方向就是我说的这一条」的强硬立场，把这块地照亮；以及一件出自这三家之一自己的材料，把这条假定顶翻。下面两节做这件事。

四·从三门外学科各取一条驱动

选这三门学问，不是因为它们与中医像，恰恰因为它们与中医不像，也彼此不像。选它们的唯一理由是：三家在同一件事上给出了互不相容的强硬答案，而且每一家都主张**只有自己那条驱动是决定性的**。凡是说「这是多种因素共同作用的结果」的立场，在这里不算立场，因为它什么也没排除。

四之一·分析化学：物种是被程序切出来的

问题的形状：一份土样里含有铅，总量可以测得很准。但铅的毒性、迁移能力、被作物吸收的可能性，取决于它以什么形态存在——是可交换态、碳酸盐结合态、铁锰氧化物结合态，还是残渣态。要知道形态，得做形态分析。

分析化学在这件事上的强硬主张是：**这些形态的显露，完全由测定路径与样品基质之间的相互作用决定**；离开一套具体的提取程序，「这份土样里有多少可交换态铅」这句话没有意义。换一套提取程序，同一份留样会给出另一份形态清单，而两份清单都不是错的。

这不是某个学者的私见，而是这门学问的正式表述。国际纯粹与应用化学联合会在关于元素形态与分级的术语指南里，明确把这类分级称作**操作定义的**：所得的「组分」由所用的程序定义，而不是由被测物的某种独立存在方式定义。特西耶等人一九七九年那套广为使用的五步顺序提取，以及后来欧洲标准局主导的三步方案，都写着同样的话。

它的动力式是这样一句：

测定路径与样品基质之间的矛盾，逼出这份样品显露成什么物种清单；清单一旦改变，回过头改写下一轮该用什么程序、以及标准物质该怎么配。

时序读数：先动的通常是基质一侧——出现了新的样品类型（污泥、生物炭、纳米材料），旧程序在新基质上给出荒唐的结果；提取程序在数年内跟上；标准物质与实验室间比对再滞后数年。跟不上的那一步永远是标准物质：认证一份新的参考物质需要多个实验室长期协作。

这一家自曝的一处：当法规限值先改——比如某项限值不再管总量而改管某个形态——程序被迫跟着改，此时是被规定的清单在驱动测定路径，而不是路径在驱动清单。分析化学知道这件事，把它写在法规遵从的章节里，不写在方法学原理里。

四之二·教育诊断学：分类只有接上一条可执行的路径才活得下来

问题的形状：一个学生读不出字来。他属于哪一类？发展性阅读障碍？语音意识薄弱？教学机会不足？工作记忆容量限制？这些分类都有测量工具，也都有拥趸。

教育诊断学里有一支的强硬主张是：**一个分类是否成立，完全取决于它能否接上一条可执行、可核查、能改变结果的教学路径；接不上路径的分类会自行消失，无论它在统计上多干净。**这一支的证据是显性的：几十年里被造出来的学习者类型学多得数不清，能活下来的极少数，全都是能直接推出「那么下一步该教什么」的那些；能力—处理交互作用的研究热潮曾经产生大量类型，绝大多数因为推不出差异化的教学而消失。干预反应模型把这条主张推到极端——它干脆不先分类，而是先施加一层标准教学，用「对教学的反应」本身来定义类别。

它的动力式：

学生当下显露出的表现与他所处的教学条件之间的矛盾，逼出教学路径改变；新的路径回过头改写了什么表现能被看见、以及条件该怎么配置。

时序读数：先动的通常是显露一侧——一次显眼的失败，一批考不出来的学生；教学路径在一到两个学年内跟上；评价制度与班级配置滞后更久，常常滞后到下一轮课程改革。

这一家自曝的一处：当某个标签一旦写进法定的安置与拨款程序，它就被冻结了——此后是行政条件在驱动教学路径，而不是学生的表现在驱动。特殊教育分类的行政固化是这一支自己反复批评的现象，他们把它当作政策病，不当作驱动方向可以翻转的证据。

四之三·文物与艺术品保存：作品的同一性由承认它的条件场决定

问题的形状：一件唐代木雕，历代修补，表面有明代的漆、清代的补手、二十世纪的加固钉。哪些部分算「这件作品」？一件行为艺术，作者已故，还能不能重演？重演的那一场是不是那件作品？

文物与艺术品保存与艺术本体论里有一支的强硬主张是：**一件东西是不是某件作品，完全由承认它的条件场决定——著录规程、展陈惯例、机构判定、共同体的接受；离开这套条件，物件只是材料。**古德曼关于亲笔性与他笔性的区分把这条说得最清楚：对于以记谱方式存在的作品，同一性由记谱系统裁定，物质载体无关；迪基的体制论则把「艺术品」直接定义为被某个体制授予候选地位的人工物。表演作品的本体论争论几十年，落点始终在「什么条件下这一场算作那部作品的一次演出」。

它的动力式：

现存实物形态与制作或演绎路径之间的矛盾，逼出这件作品赖以存在的条件场改变；新的条件场回过头改写了什么算实物、什么算路径。

时序读数：先动的通常是路径一侧——出现了新的复现手段（高精度扫描、三维打印、数字重演），旧的条件场无法裁定新造物的地位；著录与展陈规程在数年内跟上；法律与市场惯例滞后最久。

这一家自曝的一处：赝品被科学检测识破时，物件没有任何改变，而作品消失了——这一次是物质读数在驱动条件场，不是条件场在驱动。保存学把它记在技术鉴定的成就栏里，不把它记作方向翻转的证据。

四之四·三家互不相容在哪一点上

三家各自主张一条驱动，而三条驱动被驱动的那一端互不相同：化学说被驱动的是**显露出来的清单**，教育学说被驱动的是**路径**，艺术学说被驱动的是**条件场**。这不是三种角度，是三条不能同真的主张。若化学对，那么在教育学的场景里，学生的类型清单应当由测量程序与学生所处环境的相互作用决定，而不由「能否接上教法」决定；若教育学对，那么在化学的场景里，一份形态清单能否活下来应当取决于它能否接上一条可执行的处置（修复、固化、淋洗），而不取决于程序本身；若艺术学对，那么在另两个场景里，决定权应当在承认的体制手里，而不在程序或路径手里。

这三条不能靠「判谁对」化解，因为三家根本不在同一个语域里作答。它们能被化解的唯一方式，是承认一件三家都没说的事：**谁驱动谁，不是给定的，它随轮次轮转**。

要说这句话，需要材料。而按本章的纪律，这件材料必须来自三家之一自己——从外面搬一个理由来，得到的只是第四种立场，不是对争论的取消。

五·翻转发生在哪一处：分析化学自己的那件材料

材料在分析化学手里，而且是这门学问最引以为豪的一段历史。

一九七九年之后，顺序提取被大量采用，随之而来的是一个尴尬：不同实验室对同一份样品给出的形态数据无法比较。步骤顺序、提取剂浓度、液固比、振荡时间、离心转速，每一处细微差别都会改变结果，更麻烦的是提取过程中的再吸附与再分配——被从一相里提出来的金属，会在下一步之前重新附着到另一相上。这些被称为伪影，是这门学问长期的心病。

化学界的解决办法不是去找「真实形态」。他们做了另一件事：**把程序本身固定死，然后围绕这套被固定的程序，认证一批参考物质**。欧洲标准局主导的三步方案就是这样成型的，配套的认证参考物质给出的不是「这份土样里真实的可交换态含量」，而是「按本方案操作，一组合格实验室能重现的读数」。认证证书上写得明白：这些数值是操作定义的。

到这里为止，通常的叙述就结束了，结论是「标准化解决了可比性问题」。但如果把眼睛往后移一格，会看到另一件事：

在这套程序被认证之后，驱动方向翻转了。

在认证之前，是基质在驱动清单：出现新的样品类型，就得改程序。在认证之后，是被认证的清单在驱动程序与基质两侧——新的参考物质必须按这三步来配制，否则无法用于比对；新的方法要进入标准，必须先证明它在这套参考物质上给出一致读数；新一代研究者被训练成用三步方案的语言思考土壤中的金属，以至于他们提出的科学问题本身就是按这三个组分切分的。

也就是说，在这门学问自己的历史里，**同一个对象的驱动方向在一个可以指认的时点上换了一次手**。而这件事从来没有被当作关于驱动方向的证据使用，因为在分析化学自己的问题结构里，它回答的是另一个问题——「如何让不同实验室的数据可比」。方向翻没翻，与那个问题无关，所以没有人去看。

这件材料一旦被摆到这个位置上，三家的共同假定就顶不住了。它顶不住的原因不是有人反对，而是**反例出自三家中最硬气的那一家自己**：那家主张「路径与基质决定清单」的化学，自己的历史里存放着一段清单反过来决定路径与基质的时期。

于是可以说出那句被三家共有前提挡住的话：

驱动的方向不是给定的。它取决于这一轮谁先动；而这一轮被驱动的那一样发生改变之后，会回过头改写前两样，从而决定下一轮由谁先动。

三条驱动不是三支互不相干的箭头，是一个**时序循环**里的三段。分析化学看到的是「基质先动」那一段，教育学看到的是「显露先动」那一段，艺术学看到的是「路径先动」那一段。三家都没有错，三家都只看见了自己那一轮。

六·承重判断：随程物

现在可以把判断写出来。

「证」不是一个尚未被测准的隐藏实体，不是一套评分者信度不足的主观编码，也不是一种原则上不可言说的整体把握，而是一类只在某一套识别程序运行期间才具备同一性、并且会反过来改写这套程序的对象。

这类对象，本章称为**随程物**。

它的两个性质要分开写清楚，因为只有第一个性质的东西早已有名字，不必再造一个。

性质一：它在程序之外不能被独立指认。不是难以指认，是没有可指认的东西。撤掉那套识别程序，剩下的不是一个未被测准的它，而是它不存在。这句话听上去激进，但它在别的领域里是常识：撤掉三步提取方案，「可交换态铅」这个对象不存在——存在的是一份土样和其中的铅总量；撤掉一套评价工具，「阅读理解能力第三级」这个对象不存在——存在的是一个会读书或不太会读书的孩子；撤掉一套著录规程，「原作构件」这个对象不存在——存在的是一堆木头和漆。

性质二：它会回写它所依附的那套程序。这一条把随程物与单纯的操作标签分开。操作标签是死的：它由程序生成，程序改了它就换一批，它自己不参与决定下一版程序长什么样。随程物是活的：由它切分出来的世界，会规定下一轮的程序必须处理哪些东西、必须在哪里落刀、什么算作合格的读数。三步方案切出来的三个组分，反过来规定了此后二十年新方法必须报告哪三个数。

有了这两条，开头那件怪事就有了单一的来源，不必再拆成两半解释：

它不可还原，是因为每一轮还原研究在自己结束之前，已经通过修订识别程序把它自己的对象重划过一次。还原研究的标准做法是先按现行标准分组再找指标，而现行标准本身正是上一轮研究的产物——上一轮找到的指标会进入下一版诊断标准的条目，改变入组的切法。于是第二轮找到的指标之所以与第一轮不重叠，不是因为两轮里有一轮做错了，而是因为两轮切的不是同一个对象。这个过程可以无限进行下去而永不收敛，且每一轮内部都自治。

它有效，是因为它的划分与当下可执行的处置手段清单共用同一套切法。中医的证名从来不是纯粹的状态描述，它们大多同时是处置的索引：「肝郁脾虚」这个名字里同时写着「疏肝」与「健脾」两个动作；「气滞血瘀」里写着「行气」与「活血」。一位医师判定一个证，实际上做的是在当下可执行的动作库里选一条落刀线。选得准，指的不是命中了病人体内一个固定的东西，而是这一刀切出来的那一份，恰好是他手上的方药能改变的那一份。

这就解释了为什么两位医师用两套划分处置同一个病人可以都有效：他们各自的划分都与各自的动作库匹配。也解释了为什么规范化没有导致疗效塌方：规范化换掉的是划分，同时也换掉了配套的动作库（新的指南、新的成方、新的临床路径），两侧是同步换的。真正会出事的，是划分换了而动作库没跟上，或者动作库更新了而划分被冻结——本章后面会说明这两种脱钩各自会显露成什么。

为什么这三家自己到不了这一步。不是因为它们疏忽。分析化学的核心问题是可比性，它一旦承认「操作定义的组分是一类真实存在的东西」，就必须为这些组分承担本体承诺，而这会立刻毁掉它最要紧的免责声明；所以它必须停在「这些数值不代表真实物种」这句话上，这句话是它的护栏，不是它的盲点。教育诊断学的核心问题是「下一步该教什么」，它一旦承认类别只在程序内成立，就无法向家长与行政部门交代分类的正当性；停在「分类要接得上教法」这句功能性表述上，对它是最省事的。文物与艺术品保存的核心问题是同一性的裁定权，它已经把裁定权交给了体制，再往前一步就要问「体制的划分从哪里来」，那不是它的问题。一个领域看不见什么，通常不是它的缺陷，是它的问题结构的影子。

七·两根轴与四格

一个名字不是一个辨别维度。要能分东西，得有两根轴。

- 轴一：这个对象在识别程序之外能否被独立指认。两端是「能」与「不能」。
- 轴二：由它切出来的划分，是否回过头约束下一版程序。两端是「不回写」与「回写」。

	不回写程序	回写程序
程序之外可独立指认	① 实体：血糖、幽门螺杆菌、骨折线。换一台仪器、换一套标准，它还在那儿；它的存在不规定下一版仪器长什么样。	② 被规定的实体：颗粒物粒径切割点、法定血铅阈值。物质基础独立存在，但一条法规选定的切法反过来规定了此后所有仪器必须在哪里落刀。
只在程序运行期间成立	③ 操作标签：顺序提取的各组分（按其自身声明）、一份量表的分数段、一次评级的等第。程序换了就换一批，它自己不参与决定下一版程序。	④ 随程物：「证」、表现性评价里的能力类别、需要被重演的表演作品、经过多次修复的古物中的「原作部分」。它只在程序里成立，而它切出来的世界规定了下一版程序必须处理什么。

四格都要能落到具体的人和事，否则这张表只是修辞。

第①格是还原医学的主战场，也是它成功的地方。这里的成功有一个可核查的签名：跨研究的指标重叠率随年份与样本量单调上升。幽门螺杆菌从被怀疑到被确证，走的正是这条曲线。

第②格常被误认为第①格。血铅的健康限值从一百微克每升降到五十再降到更低，铅在血里的存在方式一点没变，但每一次限值下调都逼着检测方法提高灵敏度、逼着质控样品重新配制、逼着流行病学重新划分暴露组。这里有回写，但被回写的是测量，不是对象。

第③格是绝大多数评估工具自己声明的位置。这是一个谦逊的位置，也是一个安全的位置——只要声明「本量表得分不代表任何实体」，就不必为它承担本体承诺。分析化学的形态分析明确地站在这一格，并把这个声明写进认证证书。

第④格是本章要打的那一格。它与第③格的差别只有一处，就是回写；而正是这一处让它变得难以处理。

八·为什么第④格能通过全部形式审查

一个值得打的靶，必须是所有现成的检查都发现不了的那一格。若一眼就能看出它坏，那就不需要一个新框架，任何人都看得见。第④格满足三条签名。

签名一：它在短期指标上表现为改善。

规范化推进的每一步，短期指标全部向好。评分者一致性上升，术语统一，教材可编写，考试可命题，临床路径可制定，多中心研究可开展，论文数量上升，标准被国际组织采纳。没有任何一项常规指标会因为对象被重划而变差——因为常规指标全部是在**某一版程序内部**计算的，而每一版程序内部都是自治的。

同样的形势出现在教育测评（换一套工具后一致性上升、分类改变被记作版本差异）和文物著录（修订规程后著录质量上升、构件归属变更被记作勘误）。在**任何一版程序内部**，第④格与第①格看起来一模一样。

签名二：它的失效不产生任何信号。

这一条最要紧。设想一次严重的失效：某一版标准把两个在处置上应当分开的人群合并成了一个证，于是此后十年里所有按这版标准做的研究，都在一个混合人群上寻找指标，找不到，被归因于样本量不足。

这次失效会在哪里被记录？不会在任何地方。不会在临床记录里——每一个病人都被判了一个证、开了一张方、记了一个转归；不会在质控报告里——一致

性很好；不会在研究报告里——阴性结果被归因于样本量；不会在标准修订说明里——修订说明记的是「新增若干条目、调整若干权重」，不记「本次修订导致既往病例归类改变的比例」。

跨轮次的不可比性，不出现在任何一份记录的任何一个字段里。这不是有人隐瞒，是没有这个字段。

签名三：它自我加固，越正规化越严重。

程序切得越死、执行得越严，每一版内部的自治度越高，跨版的重划幅度也越大。一个松散的、允许医师个人裁量的体系，其分类边界是模糊的、渐变的，重划是缓慢渗透的；一个刚性的、条目化的体系，每一次修订都是一次整体切换，存档病例的归类会成批改变。规范化程度与本章关心的那个量是正相关的，而规范化程度同时又与所有常规质量指标正相关。

所以第④格的处境是：**做得越好，问题越大，而所有的表都显示做得越来越好。**

九·一个可以清点的读数：随程率

一个判断如果只能被同意或反对，它就还没落地。要落地，得给出一个可以由别人在别人的材料上清点出来的数。

九之一·定义与编码规则

随程率＝在最近 N 次「对象归属发生变更」的记录中，由识别程序或标准发生变更所引起、而非由对象本身出现新读数所引起的次数，占 N 的比例。

它是一个比例，无量纲，跨领域可以并排比较。清点它需要三条编码规则，缺一条这个数就不可复算：

- **规则一：什么算一次归属变更。** 同一个可追踪的个体（一位病人、一名学生、一份留样、一件构件），在两个时间点上被记入了不同的类别。仅仅是描述文字的改动不算，必须是类别标识改变。
- **规则二：什么算「对象本身出现新读数」。** 在两次归类之间，该个体的原始观察记录中出现了此前没有的项目或数值变化，且该项目是现行判定规则里的入判项。此处不要求判断这个新读数「是否足以」导致换类——只要它存

在，这一次就记为对象侧变更。这条规则是有意从宽的，它把疑难案例全部判给对手一侧。

·**规则三：什么算「程序或标准变更」**。判定规则的文本、权重、切分点、入判项清单，或执行判定的工具、人员资质要求发生了改变，且该改变在时间上早于本次归类变更。

三条规则决定了这个数的大小，所以任何一次报告随程率，都必须同时公布这三条的本地化版本。这个要求不是形式，它是本章全部结论中最容易被架空的那一环——一个不公布编码规则的随程率是没有意义的数字。

九之二·四个领域里的同一个比率

中医临床：调取一家医院近十年的电子病历，取所有「同一病人在两次就诊被判为不同证型」的记录对，逐对查当日的四诊记录，看是否出现新的入判项。分母是全部换证记录对，分子是当日无新入判项的那些。这个数在标准修订前后的年份里应当出现跳变。

土壤与沉积物形态分析：取一批留样，分别按两套顺序提取方案测定，逐个样品比对各组分归属。分母是全部样品，分子是形态清单发生改变的样品数——留样本身在两次测定之间没有变化，所以这里的随程率在原理上接近于一。这正是化学界为什么要固定程序：他们知道这个数很高，只是没有把它当作一个数来看。

教育测评：一个学区更换评估工具后，重新评估的学生中被归入不同能力类别的比例，扣除其间行为记录出现实质变化的那些。这个数在几项已发表的表现性评价概化研究里可以间接推算，但从未被单独报告过。

文物著录：一家馆修订著录规程后，被重新判定归属的构件数中，其间没有新的检测报告支持的那部分。这个数在修订说明里通常一字不提。

四处量纲完全不同——病历条数、样品数、学生数、构件数——而比率的定义是同一个。

九之三·与既有主要指标正交的那个例子

一个新造的读数如果与既有的主要指标高度相关，它会立刻被读成那个指标的一个代理，白造。所以必须举出一个「既有指标最好看、而这个数最难看」的例子，举不出来就该重造。

例子是这样的：某病种的证候诊断标准修订后，专家间一致性系数由零点六二升至零点八五，达到该领域公认的优秀水平；量表的重测信度、内部一致性、结构效度指标全部改善；标准被行业采纳，进入临床路径。与此同时，把该院近五年的存档病例按新标准回溯判定，四成病例换了证型，其中绝大多数病例的原始四诊记录里没有任何新的入判项——他们的随程率接近零点九。

一致性最高的那一版标准，随程率也最高，因为一致性正是靠把切法固定得更死换来的，而切法越死，换版时的整体位移越大。这两个数在方向上是相反的，任何一个现行质量框架都不会同时报告它们。

九之四·这个数的伦理边界

随程率是一个比率，比率会被拿去考核人。所以有三条必须写死，缺一条本章的这一部分不应被采用：

其一，这个读数指的是位置，不是人。高随程率描述的是「这一类对象与这套程序的关系」，不描述任何一位医师、教师、检测员的水平。一个把病人按现行标准准确判定的医师，其经手病例的随程率可以很高，那与他的判断质量无关。

其二，不得为了把这个数做好看而冻结标准。在任何以安全为要的系统里，都不得以「降低随程率」为由推迟本应进行的标准修订。这一条与下文的反噬预言直接相关。

其三，已经按这个读数做出不利安排的，必须公开编码规则并给出复核通道。由于随程率的大小完全取决于三条编码规则的本地化写法，任何据此作出的评价，都必须允许被评价者按公开规则复算，并对编码争议给出裁决路径。

十·一句不含情态词的判据

一个判断要能被用，得配一句可以拿去问记录、而不是拿去问人的判断的问话。这句问话里不能有「应当」「充分」「真正」「合理」这类词，因为带这些词的问话只能得到意见。

本章的判据是：

最近 N 次分类变更中，有几次的当日记录里没有出现新的观察项？

把它拿到五个场景里跑一遍，五个答案互不相同：

1. **一家中医院的门诊系统：**查最近两百次证型变更记录对，逐对调取当日四诊记录，统计其中当日无新增入判项的次数。答案是一个介于零和二百之间的数，且这个数在标准修订当年会显著跳高。
2. **一间土壤实验室的留样柜：**同一批留样在两套提取方案下的形态归属改变数。由于留样在两次测定之间没有任何新读数产生，答案在原理上等于全部改变数——这是一个天花板案例。
3. **一个学区的特殊教育安置记录：**换用新评估工具后被改变安置类别的学生中，其间行为观察记录未出现新条目的人数。这个答案通常介于中间，因为换工具往往伴随着重新观察。
4. **一家美术馆的著录档案：**修订著录规程后由「原作构件」改判为「后加」的构件中，其间未更新检测报告的数量。这个答案往往接近全部——规程修订通常不伴随重新检测。
5. **一家医院的检验科：**某项指标参考区间调整后，由「异常」变为「正常」的既往报告数中，标本未复检的比例。这是第②格的案例，答案很高但性质不同——被重划的是判读阈值，不是对象。

五个答案里，至少有两个不能从本章的判断直接推出来：土壤留样在两套方案下的实际重叠比例是多少，以及学区换工具时有多大比例伴随了重新观察——这两个数只能查，不能想。这一点很要紧：如果五个答案全都能从命题直接推出，那它就不是判据，只是命题的复述。

十一·在别的领域里能预测出什么

类比是「像」，兑现是「在那个领域里也能据此预测出一件具体的事」。以下十二处，每一处都给出一个可以去查的预测。

一、分析化学：新的提取方案被采纳后的头五年，围绕该方案发表的研究中，被作为科学问题提出的形态类别，其构成应当向新方案的组分划分靠拢，而不是保持上一版的划分。可查方式是对该领域文献的标题与关键词做一次分期词频比较。

二、环境法规：当某项限值由管总量改为管某一形态之后，随后五年内新增的检测方法标准，其步骤设计应当以能否分出该形态为首要约束，即便这一约束在分析上并不是最优的。

三、教育测评：同一批学生在两套高信度工具下系统性换类的情形，不应当被读作至少一套工具无效。本章预测：两套工具各自配套的干预在各自的分类下都能显示效果，而交叉配对（按甲工具分类、用乙工具配套的干预）效果显著下降。这是一个可以做的实验。

四、干预反应模型：以「对教学的反应」定义类别的体系，其随程率应当显著低于以「学生特质」定义类别的体系——因为前者把处置手段直接写进了分类的定义里，两侧不会脱钩。

五、精神医学诊断：诊断手册每一次大改版后，同一批存档病例的回溯归类改变比例，应当与该次改版中「诊断条目与可报销治疗方案对应关系」的改动幅度正相关，而与该期间神经生物学发现的数量相关性很弱。这是一个可以用已发表版本比较研究做的检验。

六、疾病编码：国际疾病分类版本更替时，各章节的归类迁移率应当在「治疗手段近年变动大」的章节里最高，在「治疗手段稳定」的章节里最低，而与该章节所涉疾病的病理机制研究热度关系微弱。

七、文物与艺术品保存：当一种新的无损检测手段普及之后，著录规程的修订应当滞后于该手段普及三到七年，而在修订完成之后，此后被提出的鉴定问题应当按新规程的分类重新组织。

八、表演作品重演：一部依赖口传的作品，其「授权重演」的判定标准修订之后，此前被判为不合格的重演版本中应有相当比例被重新接纳，且这些版本自身没有任何改变。

九、司法证据：某类证据的采信标准修订后，重新审查的旧案中改变结论的比例，应当与该期间该类证据检测技术的进展关系微弱，而与采信标准文本的改动幅度直接相关。

十、组织绩效：一家公司更换绩效评估体系后，被重新归入不同绩效等级的员工中，其间产出记录未变化的比例，应当随新体系的条目化程度上升而上升。条目化程度越高，等级评定的一致性越好，而这个比例也越高。

十一、生态学中的物种界定：以分子标记划界的类群，其种数在标记体系更替时的变动幅度，应当远大于以生殖隔离划界的类群，而这与实际生物多样性的变化无关。分类学界称之为分类学膨胀，本章预测其幅度与「保护资金的分配是否按种数计」正相关——因为那意味着分类切法与处置手段挂上了钩。

十二、临床试验入组：一项新药试验的入组标准若以证候诊断标准为依据，则该标准在试验期间修订的话，入组人群的构成会漂移；本章预测这种漂移在长周期试验中的发生率，高于研究者的普遍估计，而现行的试验注册与报告规范里没有任何字段记录它。

十二·三处反过来给本判断加的约束

本章的判断不是从三门学问里取长补短得来的，所以它也得接受这三门学问反过来的削减。以下三处，每一处都真的削掉了本章原本打算主张的东西。

十二之一·分析化学削掉了「凡有效即随程」

形态分析里存在一类对象，是可以在程序之外被独立指认的：用同步辐射 X 射线吸收谱、单晶衍射直接看到的配位环境，不随提取方案而改变。它们是真的第①格。

如果没有这一处，本章原本会写下的更强的一句是：「凡是在复杂系统里持续有效的临床对象，都是随程物；还原主义在医学里整体失效。」这句话现在不能写了。血糖是第①格，幽门螺杆菌是第①格，骨折线是第①格，它们在医学里同样持续有效，而且完全可还原。

本章的适用范围因此被砍掉一大块，缩窄为：**只有当识别程序与处置手段共用同一套划分时，对象才是随程物。**判定方式很直接——问一句「这个分类的名字里有没有写着一个动作」。「肝郁脾虚」里有动作，「血糖七点八」里没有。没有动作的那些，不在本章管辖范围内。

十二之二·教育诊断学削掉了「追求跨程序稳定是浪费」

阅读障碍研究里有一个跨工具相对稳定的成分：语音解码的核心困难，在不同测量体系下都能被识别出来，而且干预的有效性明显依赖于抓住这个成分。它不是完全随程序漂移的。

如果没有这一处，本章原本会写下的更强的一句是：「追求跨程序稳定的分类，是把力气花在一个不存在的目标上。」这句话现在也不能写了，而且被削掉的不只是一句主张，是本章的价值排序。

被迫承认的是：**随程不是一个二分，是一个连续谱。**一个对象可以有一部分成分跨程序稳定，另一部分随程序漂移，而随程率正是这个谱上的位置读数。第

④格因此不是一个抽屉，是这个谱的一端。这一改动的直接后果是：本章不能主张「不要去找证的物质基础」，只能主张「找到的那一部分会把对象切走一块，剩下的那一块仍然是随程的，而这个搬运过程目前不被记录」。这比原来的主张弱，也更可能是对的。

十二之三·文物与艺术品保存削掉了「承认即一切」

赝品案例证明存在一个可以否决体制的物质事实层：颜料里的钛白粉不可能出现在十七世纪，无论多少专家和多少机构承认过这幅画。物质证据可以单方面推翻承认，而承认不能推翻物质证据。

如果没有这一处，本章原本会写下的更强的一句是：「对象的身份完全由划分它的程序决定。」不能写。存在一个下限：**随程物的重划有边界，边界由不可被程序改动的那些事实划定**。一位病人有没有发热，是可以在任何一版标准之外被独立确认的；能被重划的是「发热在这一版里算不算入判项、算多少权重、与哪些项组合成哪个证名」，不是发热本身。

这一处削减很重要，因为它挡住了本章最容易滑向的那个方向——一旦说成「一切都是程序造的」，这个判断就变成了不可证伪的立场，而不是可用的工具。

十三·与最近的几种说法的分界

本章的判断有一个明显的危险：它听上去像好几种已经存在几十年的说法。如果不能逐一说清差在哪里、并且给出一个能分出高下的观察，那么它就只是一次改名。以下逐一处理，每一位都写出「看到什么就说明本章对而它错，看到什么就说明反之」。

布里奇曼的操作主义（Bridgman, *The Logic of Modern Physics*, 1927, 物理学）。它说到哪一步：一个概念的意义就是相应的一组操作，脱离操作的概念没有意义。这是本章形式上最像的一位。分离线在于：操作主义**取消对象**——它主张概念之外无物，因此它对「对象反过来约束程序」这件事没有位置安放，在它看来程序的选择是自由的、外在的。本章主张对象在程序运行期间是实有的，并且会限制下一版程序的可选项。判定形式：**若观察到一套程序被采纳之后，后续程序的可选空间被它切出的划分实质性收窄**（新方法必须报告这几个

组分才被接受），则本章对而操作主义错；若观察到程序可以任意更换而不留下任何对后续的约束，则反之。分析化学三步方案的二十年历史属于前者。

克隆巴赫与米尔的构念效度（Cronbach & Meehl, 1955; Messick, 1989, 心理测量学）。它说到哪一步：构念是理论网络中的节点，测量向它趋近，效度是证据的累积。分离线在于：构念效度理论把「换一套工具、系统性换一批类别」读作**效度不足**，即测量没有对准同一个构念；本章把它读作**对象的性质**。判定形式：若同一批个案在两套各自高信度的工具下系统性换类，而两套分类下各自配套的处置在各自体系内都显示效果，则本章对而**构念效度错**——因为它预测至少有一套工具没测到那个构念，因而其配套处置应当无效。这一条是可以做实验检验的，见第十一节第三处。

波兰尼的默会之知（Polanyi, *Personal Knowledge*, 1958; *The Tacit Dimension*, 1966, 认识论）。它说到哪一步：我们知道的比我们能说出的多，技艺性知识在形式化中流失。分离线在于：默会说预测**形式化会带来有效性的损失**；本章预测**形式化不损失有效性，只改变对象边界**——旧程序绑定的那一份随旧程序退场，新程序会长出自己的一份。判定形式：若**证候诊断标准修订并被普遍执行之后，按新标准培训的一代医师的疗效不低于旧标准，而个案归类大面积改变**，则本章对而默会说错；若**疗效随规范化程度上升而系统性下降**，则反之。

法夸尔对中医临床的实践论描述（Farquhar, *Knowing Practice*, 1994, 医学人类学）。它说到哪一步：中医的知识不是命题的集合，而是在临床相遇中被实作出来的；教科书化的辨证论治是一种现代的重构。这是本领域内与本章最近的一位。分离线在于：法夸尔关心的是**知识的存在方式**（知识长在实践里），本章关心的是**对象的存在方式**（对象与程序共存亡并回写程序）。两者的经验后果不同：实践论倾向于把标准化读作对真知识的侵蚀，本章读作对象的一次搬家。判定形式：若在高度标准化的诊疗环境里，**随程率很高而疗效不降**，则本章对；若**标准化程度与疗效之间存在负向的剂量—反应关系**，则实践论对。

哈金的回路种类（Hacking, “The looping effects of human kinds”, 1995; “Kinds of people: moving targets”, 2007, 科学哲学）。它说到哪一步：关于人的分类与被分类的人互相塑造，被分类者因为知道自己被这样分类而改变，于是分类成为移动的靶子；哈金明确地把这限定在**互动种类**即人的种类上，自然种类不回路。这是本章最危险的一位近邻，因为「移动的靶子」这个说法几乎就是本章的开头一句。分离线在于回路发生在哪里：哈金的回路经由**被分**

类者的自我理解，本章的回路经由程序与对象清单之间，完全不需要被分类者知道任何事情。判定形式：若在没有自我意识的对象上——一份土样、一层颜料——也观察到程序与清单之间的双向改写（随程率高且存在回写），则本章对，而回路效应的边界条件被证明画错了；若这种双向改写只在人类种类中出现，则哈金对，本章不过是回路效应在医学中的一个应用。顺序提取的二十年史是前者的直接证据。

莱因贝格尔的认识之物（Rheinberger, *Toward a History of Epistemic Things*, 1997, 科学史）。它说到哪一步：实验系统产生认识之物，认识之物一旦被稳定下来就转化为技术对象，技术对象再去生成新的认识之物。这里有回写，也有循环，与本章极近。分离线在于方向：莱因贝格尔的循环是收敛的——对象最终被固定为技术物，退出认识前沿；本章预测**不收敛**，因为随程物的划分与处置手段库共用一套切法，只要动作库更新，它就被重新推回未定状态。判定形式：若某类对象在数十年内反复被固定又反复被重划，且每一次重划在时间上紧跟处置手段库的变动、而不紧跟检测技术的进展，则本章对而转化方向的说法**错**；若重划次数随时间递减、对象逐步沉淀为稳定的技术对象，则反之。这是本章全部可裁决主张里最要紧的一条，第十六节会把它写成正式的证伪条件。

默顿与塞尔兹尼克的目标置换（Merton, 1940; Selznick, 1949, 组织社会学）。把本章的命题剥掉全部医学名词，只剩下「一套程序与它要处理的对象互相改写，导致对象的身份随程序轮转」。这个纯形式的结构在组织研究里最像的模板是目标置换——手段取代目的。必须点名它，因为它已有八十年历史，任何读者都可能把本章读成它的一个变体。分离线在于**结果的方向**：目标置换预测**有效性下降**，手段压过目的之后目的落空；本章预测**有效性可以维持甚至提高**，而对象换了。判定形式：若程序标准化推进的同时处置有效性维持不变、而对象归类大面积改变，则本章对而目标置换**错**；若有效性随程序刚性上升而系统性下降，则这不过是一次目标置换，本章没有新东西。同族的几个反转模板——成功陷阱、能力刚性、内卷化——共享同一条预测（越成功越失败），因而共享同一条判定形式：它们全部预测**绩效下滑**，本章**不预测绩效下滑**。

操作定义的形态分析本身（Tessier et al., 1979; Templeton et al., IUPAC, 2000, 分析化学）。这一位不是内容上的近邻，是**方法上的**——它已经有一整套「用程序定义对象」的成熟做法，几十年，成千上万篇文献。若本章只是把它搬到医学里，那本章就是被它覆盖的。分离线在于本体承诺：形态分析明

确声明这些组分不代表真实物种，因而它拒绝给对象以本体地位，也就不必解释回写；本章给它本体地位，并要求把回写记账。判定形式：若一套形态划分的变更被观察到改写了后续的法规文本、参考物质配制与研究问题的提法，则这些组分不只是标签，本章对；若它们始终只是报告格式而不产生任何对后续的约束，则形态分析的自我声明是准确的，本章多此一举。

概化理论与任务抽样变异（Cronbach et al., 1972; Shavelson, Baxter & Gao, 1993, 教育测量）。它说到哪一步：表现性评价中最大的方差成分往往是人与任务的交互，换一批任务就换一批排名，因此需要大量任务才能得到可靠的概化系数。它把这件事完整地测出来了，并把它记为测量误差的一个来源。分离线在于归属：本章把这同一个方差成分读作对象随任务重划的直接证据。判定形式：若在人—任务交互方差很大的同时，各自任务集下配套的教学干预都能显示效果，则这部分方差不是误差而是对象，本章对；若只有其中一套能显示效果，则它确实是误差，概化理论对。

证候客观化研究纲领（沈自尹自一九五九年起的肾阳虚研究、陈可冀等的血瘀证研究，以及此后的证素辨证与多组学路线，中医学）。它说到哪一步：证有物质基础，只是层次多、变量密，需要更高维的技术与更大的队列。分离线在于对收敛的预测：客观化纲领预测指标集合随技术世代与样本量而收敛；本章预测不收敛，且在每次标准修订后出现重置。判定形式：若把历年已发表的同一证型标志物集合按发表年份排序，其两两重叠率随年份与样本量单调上升，则客观化纲领对而本章错；若重叠率呈现锯齿——在每次诊断标准修订后下跌，然后在下一次修订前缓慢回升——则本章对。这条检验的全部数据在既有文献里已经存在，做一次元分析即可，不需要新的实验。

最近的那一位是谁。是哈金的回路种类。它与本章共享「靶子在移动」这个核心图像，而且比本章更早、更有名。本章之所以仍不能被它吸收，只有一个理由：它的回路必须经过被分类者的心智，本章的回路不必。一份留在柜子里的土样不知道自己被分类，它照样在两套程序之间换类，而且这次换类会改写下一版程序。如果将来有人证明这类无心智对象上的双向改写其实并不存在，那么本章应当被撤回，回路种类足够用了。

十四·与本系列各章的分界

本系列近期几篇处理的是相邻的题目，其中三篇用的是同样这三门学问。它们是本章最近的邻居，读者一眼就能认出，所以必须逐篇说清差在哪、并各给一条能分开的预测。

《撑平》（《治未病立账学》第一章）（何谓未病）。那一篇说的是：一个状态的净读数持平，可以是无人干预的空平，也可以是被持续的外力撑住的撑平，而被撑的那份维持工作在账本上不设字段。它的读数是持率。差别在于被漏掉的东西不同：那一篇漏掉的是维持工作，本章漏掉的是对象的重划。同一位病人可以同时落在两篇的管辖里，但两篇问的不是同一个问题——那一篇问「这个平是谁在撑」，本章问「这个证是哪一套程序切出来的」。判决性对照：取一位他持率为零（无人在撑，各项自主维持）而随程率极高（换一版标准即换证型）的病人，《撑平》判定此人属于未病，本章判定此人的证仍是随程物且按现行标准的处置仍然有效。两篇在这一例上给出不同结论。

《无事对价》（《治未病立账学》第七章）（治未病为何实践最少）。那一篇说的是：预防兑换的是非事件，而账本只能记事件，因此一次成功的预防不产生下一轮的资源，回写为零。它的读数是阙前占比。本章与它同样处理「回写」这个环节，但方向相反：那一篇讲的是回写为零导致实践萎缩，本章讲的是回写照常发生却没有被记账导致对象漂移。判决性对照：假设某地按《无事对价》的处方设立了以反事实估计为结算对象的预防科目，而证候诊断标准仍每十年重置一次分类。《无事对价》预测预防实践量将上升；本章预测在标准重置的年份，基于旧分类累积的预防经验会大规模失效，实践量出现回落。两篇对同一项政策给出不同的时间曲线。

《代持的病》（《治未病立账学》第四章）（未病如何被识别、无症状时如何正当干预）。那一篇处理的是识别与干预的正当性，本章处理的是被识别之物的存在方式。判决性对照：一个无症状而被判定需要干预的人，其判定依据若在下一版标准里被移出入判项，那一篇的问题是「此前的干预是否仍然正当」，本章的问题是「此前被干预的那个对象现在是否还存在」。前者要一个伦理裁决，后者要一份归类变更清单——两者需要的材料不同，因而不可互相替代。

《临界回落》（本系列单篇）（为何课堂上真正的学习很少发生）。那一篇与本章只共用两门学问，处理的对象是一次未完成的转化：抵达了临界而没有被接住，于是被记为从未开始，读数是回落率。本章处理的是一类随程序存亡的实体。判决性对照：一个回落率高而随程率为零（分类体系数十年未动）的教学或

临床场景，《临界回落》预测学习或疗效低下，本章预测有效性不受本章所述机制影响。反过来，一个回落率为零而随程率极高的场景，本章预测跨轮次的经验无法累积，那一篇不作预测。

四篇合起来的关系是互补而不是重复：一篇记漏了维持工作，一篇记漏了非事件回报，一篇记漏了无症状者的处置正当性，本章记漏了对象本身的重划。**四种漏记可以同时发生，而它们叠加时的后果不是相加**——当维持工作不被记录、且对象每隔十年被重划一次时，跨轮次的经验累积速度不是被减慢，是被清零，因为上一轮的记录连指称对象都对不上。

十五·本章不适用的地方

一个判断如果到处都适用，它多半什么也没说。以下是本章明确不管的地方。

第一，程序之外可独立指认的对象。血糖、病原体、骨折、肿瘤大小。这些是第①格，还原对它们完全有效，本章的机制不适用。若有人用本章去为「拒绝检测」辩护，那是误用。

第二，识别程序与处置手段不共用一套划分的场合。本章全部的解释力来自这两侧的耦合。在两侧解耦的地方——比如一套纯粹描述性的分类体系，与之相配的处置手段库并不随之更新——对象不会被处置手段拉着重划，本章的机制不发动。

第三，单次诊疗内部。本章说的是跨轮次的现象，最小的观察单位是一次标准修订。用它去评价某一位医师某一次判断的对错，是尺度错用。

第四，随程谱上靠近稳定一端的那些成分。如第十二之二节所述，随程不是二分。一个对象里跨程序稳定的那部分，仍然应当被当作可还原的对象去研究，而且那正是还原研究应当集中火力的地方。本章不主张停止寻找物质基础，本章主张在找到一部分之后要记录被搬走的是哪一块。

第五，本章管不了的一个洞：随程物是怎么诞生的。本章说清了它一旦存在会怎样运转，但没有说清一个新的随程物是如何第一次被切出来的——是先有了一批新的处置手段，还是先有了一次识别上的困难？本章手上的材料不足以回答，姑且把它交出去。

十六·怎样才算本章错了

十六之一·先写出最强的那个竞争解释

不挑好打的靶子。本章面对的最强竞争解释是这一句，它最朴素，也最难反驳：

其实不就是标准还不成熟吗？定义模糊、样本不够、技术不行，等它成熟了自然就收敛了。

这句话能解释本章引用的几乎全部现象：跨研究重叠率低（样本不足）、换版换类（定义在改进）、一致性上升而重叠率不升（改进还没到位）。它不需要任何新概念。要证明本章不是它的一个花哨说法，必须拎出一个**只有本章的机制成立才会出现、成熟度假说预测不出的变量**，并在控制掉成熟度之后仍然要求它成立。

那个变量是**重划与什么同步**。成熟度假说预测：分类的重大变动应当跟随认识的进展，即跟随病理机制研究与检测技术的进展。本章预测：分类的重大变动跟随**处置手段库**的变动——新治法、新方药进入目录、新的临床路径成型——而与病理机制的进展关系微弱。

这两个预测在同一份材料上给出相反的读数，因此它是可裁决的。

十六之二·正式的证伪条件

条件一（主条件，机制证伪）：把一九四九年以来六次以上全国性证候分类标准的重大修订逐次编年，对每一次修订，分别统计其前五年内（甲）新治法或新方药进入国家级目录、指南、临床路径的数量，与（乙）该证型相关病理机制的高影响力发表数量。**控制住技术世代与研究总量之后，若修订时点与甲的变动之间不存在时间上的领先关系，或甲与乙对修订幅度的解释力没有显著差别，则本章的核心机制为伪——届时应当把不收敛归因于标准的不成熟，而不是归因于对象的随程性。**

条件二（顺序，轮次形态）：本章预测的不是一次先后，是一个循环。完整形状是：**处置手段库先动 → 分类在若干年内跟上 → 新分类回写下一轮临床试验的入组标准与考核口径 → 下一轮的发起处换到监管或支付一侧**。若在六次以上的修订中，发起处始终是同一侧（永远是学术研究先动，或永远是行政先动），则本章关于「发起权轮转」的部分为伪，本章将退化为一条单向的因果说法。

条件三（重叠率的形状）：把历年同一证型的标志物研究按发表年份排序，计算相邻研究间的指标集合重叠率。**成熟度假说预测单调上升，本章预测锯齿——每次标准修订后下跌，修订间隔内缓慢回升。若曲线呈单调上升，本章错。**这一条最便宜：所需数据全部在已发表文献里，一次系统综述即可完成，不需要新的病人、新的样本、新的伦理审批。

条件四（正交性）：若在多家机构的数据中，随程率与评分者一致性系数呈显著正相关的关系不成立——即一致性最好的机构随程率并不更高——则本章关于「规范化自我加固」的那部分为伪。

条件五（本体地位）：若在无自我意识的对象上（土样、颜料层、构件）观察不到程序与清单之间的双向改写，即操作定义的划分从不回写后续程序，则本章提出的第④格与第③格没有实质差别，随程物这个范畴应当被撤销，回到操作标签即可。

条件六（有效性脱钩）：本章主张有效性来自识别程序与处置手段库共用一套切法。若观察到某一版分类与处置手段库明显脱钩（分类修订而动作库未更新，或反之）而疗效不受影响，则本章关于有效性来源的解释为伪。

条件七（交叉配对）：按甲体系分类、施以乙体系配套的处置，其效果应当显著低于同体系配对。若交叉配对与同体系配对的效果没有差别，则本章错——那说明处置的效果与分类无关，分类只是装饰。

条件八（编码可复现）：若两组独立编码者按本章第九之一节的三条规则对同一批记录计算随程率，所得结果的一致性系数低于该领域可接受的下限，则随程率不是一个可用的读数，本章的操作部分作废，判断部分需要另找读数。

十六之三·样本纪律

条件一涉及跨案例比较，因此必须写死几条，否则它只是一个说法：

- **样本量不少于六个单元**，且必须在同一制度环境内取。可用的单元是：同一部委下辖的多个病种证候标准、同一学会的多次修订版本、同一省份的多家三甲医院。
- **严禁用两个国家或两个地区充数。**「甲国如此、乙国如彼」这种对比在文化、规模、监管、支付方式上处处不同，所谓「控制住某个变量」在方法上站不住。它可以用作启发式的例证，**不能充当证伪检验。**

- 若确实只能作跨国比较，必须在正文中写明这是启发式例证，并另给一个同质多单元的设计顶上。

十六之四·本章可以被分层引用

本章的主张分三层，各自可以独立倒下：

- **第一层（最强、最易倒）**：随程物作为一个本体范畴，即第④格与第③格有实质差别。它被条件五直接检验。
- **第二层（分析工具）**：两根轴与四格。即便第一层被推翻，这张表仍可作为一个描述性的分类工具使用，用以区分不同对象与其识别程序的关系。
- **第三层（最稳、最便宜）**：一条不依赖前两层的经验主张——**同一证型的标志物研究，其指标集合的跨研究重叠率不随年份与样本量单调上升，而在诊断标准修订后出现重置**。这一条可以用现成文献检验，若它成立，则无论随程物这个范畴是否被接受，「继续加大样本量就会收敛」这个政策判断都需要重新考虑。

说明本章可以被分层引用，是诚实，也是保护。

十六之五·如果本章被证伪，我们会学到什么

若条件一失败——分类修订确实紧跟病理机制进展——那么七十年不收敛就必须由别的机制解释，而最可能的解释是分类的粒度选错了：对象是可还原的，只是被切成了错误的大小。那会指向一个完全不同的干预方向：不是去记录重划，而是去重新选择切分的层次。这个结果本身是有价值的，它会把资源从「记账」转向「重切」。

十七·这个判断被采用之后会怎样反噬

任何一条判断，一旦被制度采用，最省事的执行方式往往恰好摧毁它要保护的的东西。本章的反噬是清楚的，而且我看不到出口。

反噬一：把随程率写进考核，最省事的达标办法是冻结标准。一个机构若被要求降低随程率，它可以不去改进任何实质的东西，只要推迟标准修订即可——不修订就没有程序变更，随程率自然为零。而标准长期不修订，恰恰会让分类与当下可用的处置手段脱钩，摧毁本章所说的有效性来源。**降低随程率的最快路径，正是消灭有效性。**

反噬二：要求公布归类改变比例，最省事的达标办法是把编码规则写松。由于随程率完全取决于三条编码规则的本地化写法，任何机构都可以通过把「新读数」定义得极宽（任何一次问诊记录的文字变动都算新读数）来把这个数压到很低。规则一旦成为考核依据，规则本身就成了被优化的对象。

反噬三：把「对象会被重划」这件事讲出来，最省事的接收方式是虚无主义。「反正都是程序造的」这句话可以被用来拒绝一切标准化努力，也可以被用来拒绝一切还原研究。第十二之一与十二之三节的两处削减正是为了挡住这个方向，但削减写在论文里，口号传得比论文快。

出口在哪里？我没有找到。任何鼓励标准稳定的考核都会切断分类与动作库的同步；任何鼓励同步的考核都会让随程率变得不可解释；任何要求公布编码规则的制度都会让规则本身成为博弈对象。本章能做的只有一件事：**把这三条写在这里，让采用者知道自己在选哪一种坏。**

伦理边界另有两条，与第九之四节的三条并列：其一，本章的任何部分都不得被用作延迟或拒绝有效治疗的理由——「反正对象会被重划」不构成不处置的依据。其二，在按旧标准接受过处置的病人身上，标准修订不应导致其既往处置被追溯性地判为错误；处置的正当性以处置发生时的现行标准为准，这一条应当写进标准修订的过渡条款。

十八·本章自己落在哪一格

一条关于「对象随程序重划」的判断，如果不能用在自己身上，它就是自封的。

本章提出的随程率，按本章自己的两根轴判定，落在第④格。它在程序之外不能被独立指认——撤掉第九之一节那三条编码规则，「随程率」这个对象不存在，存在的只是一堆归类变更记录。它也会回写：一旦被采用，各机构为了让数字好看，会去修改「什么算一次程序变更」的本地定义，而这些修改会成为下一版编码规则的既成事实。

这个定位有三个具体后果，本章必须认下来。

后果一：本章不能靠被引用来证明自己。引用量上升意味着更多机构按各自的本地规则计算随程率，而本地规则的分化恰恰会提高本章这个对象自身的随程率。**本章越成功，本章自己越随程。**

后果二：本章给自己设了一条十年后的自检线。如果到二〇三六年，没有任何一组独立研究者能按本章公布的三条规则，在同一批病例上复算出彼此一致的随程率，那么按本章自己的第八条证伪条件，随程率就是一个第③格的操作标签，而不是第④格的读数。届时本章的第一层应当被撤回，只保留第二层的分类工具与第三层的经验主张。

后果三：本章所处的位置正是它所描述的那一格。学术评价体系本身就是一套识别程序，「有创新的论文」是它切出来的对象，而这个对象与当下可执行的处置手段库（期刊栏目、基金方向、评审口径）共用一套切法。本章若被接受，多半不是因为它对，而是因为它恰好落在当下这一版切法能够识别的位置上；本章若被拒绝，也未必是因为它错。**这句话对本章不利，但按本章自己的判断，它必定成立。**

十九·结论，与一条写死日期的赌注

七十年的还原研究没有失败在技术上。它每一轮都做对了自己那一轮的事，只是每一轮结束之前，它要还原的那个对象已经被这一轮自己参与修订的识别程序重划过一次。不收敛不是精度问题，是本体问题；有效不是运气，是因为切法与手上的动作库共用同一把刀。

这两件事有同一个来源，因此不必分头解释。承认这一点的代价是承认一类对象的存在：它只在程序运行期间成立，并且会反过来改写这套程序。**承认它，比继续把它当作噪声要贵，因为要为其开一个字段；不承认它，代价是每十年把上一轮的经验作废一次，而这笔账目前谁也没在记。**

给出一条写死日期的赌注：

到二〇三三年十二月三十一日，至少有一部国家级或行业级的中医证候诊断标准，在其修订说明的正式文本中载明「本次修订导致的既往病例归类改变比例」这一项数据，并同时公布新旧标准之间可回溯的归类映射规则。

什么不算命中，一并写死：

1. 该数据只出现在学术论文或会议报告中，未进入标准修订的正式文本，不算命中。
2. 修订说明只写「本次修订与旧版存在若干差异」而不给出比例数字，不算命中。

3. 给出了比例但不公布可回溯的映射规则（即第三方无法把旧病例按旧标准复现归类），不算命中。
4. 由制定方自行说明而无任何可供第三方复算的数据集或抽样样本，不算命中。
5. 该项数据由本章作者或其直接合作方参与推动写入的，不算命中——赌的是这件事会自己发生，不是赌它能被推动。

若到期未兑现，本章关于「这个字段的缺失是结构性的、不是疏忽」的判断得到反面的证据：那说明这不是一个不被记账的位置，而只是一件还没轮到去做的事，本章对其结构性的估计过高了。

注释

〔一〕本章所称「识别程序」，指从原始观察到类别判定之间的全部成文规则与执行安排，包括入判项清单、权重、切分点、判定者资质与工具。它不包括处置本身。

〔二〕本章所称「处置手段库」，指在某一时点上可被合法、可及地执行的干预动作集合，包括成方、治法、器械、路径与支付覆盖范围。它的边界由可及性决定，不由知识决定。

〔三〕第九节的随程率与第十一节第十一处提到的分类学膨胀在算法上相近，但两者的分母不同：前者以个体的归属变更为分母，后者以类群数为分母。两者不可直接比较。

〔四〕第十三节中对各既有说法的表述力求公允，未作简化处理；若有表述失当，责任在本章。

〔五〕第十六节条件一所需的编年材料，目前尚无公开的整理版本，这是本章没有自己完成该项检验的原因，也是本章把它列为第一条证伪条件而非已完成证据的原因。

第八章 消参

——一个系统可以被可靠地改变，而它的内部量永远不能被单独确定：辨识与控制是两条互不隶属的通道

提 要 「证」自二十世纪五十年代进入实验室以来，七十年间跨越六七代技术，始终没有收敛为一组稳定的、跨研究可复现的指标；与此同时，以辨证为前提的处置持续有效。通行的解释把两者拆成两件事：不可还原归因于样本量、异质性与技术精度，有效归因于药理活性、心理效应或经验规律。本章主张这个拆法本身错了。**可辨识性与可控性是系统的两个互相独立的结构性质**：一个系统完全可以在给定的观测结构下永远不可辨识，同时被可靠地驱动到目标状态。七十年的还原研究一直在用辨识的合格线去评价一个走通了控制通道的体系，而这条合格线是先验地过不去的——**结构不可辨识与数据量无关，加多少样本都不会收敛**。本章取物理学的标度流与相关变量判定、心理学的参照条件与测量不变性、工程学的辨识路径与激励设计三条互不相容的单因驱动主张相互顶撞，用工程学自己在自适应控制里已经掌握、却从未往这个方向使用的一件事实——**参数估计不收敛而闭环跟踪误差仍趋零**——推翻三家共同假定的「辨识在前、控制在后，方向固定」。由此得到一类在三家账本上都没有本体地位的量：**在辨识中被并入某个可辨识组合、因而失去单独身份，却在控制律里单独起作用的那一份**，本章称之为**消参**。给出一个跨领域量纲相同的读数**消参率**：原始参数数与可唯一确定的参数组合数之差，占原始参数数的比例；它是纯结构计算，不需要任何新数据。给出一句不含情态词的判据：**在这一次判定所依据的全部观测项里，有多少对在过去 N 次记录中始终同时出现、同时缺失？**全文最反直觉的一条裁决主张是：**规范化在结构上就是激励削减，因此越规范化，结构可辨识性越差，而每一版内部的一致性越高**。

一·引言：一条从来没有被检查过的前提

七十年的证候客观化研究，从尿十七羟皮质类固醇开始，经过免疫、微循环、自由基、基因芯片、蛋白质组、代谢组、菌群，直到多组学与机器学习，形成了一个稳定的形势：每一代都有阳性发现，而跨代之间、乃至同代不同实验室之间的指标集合重叠很少，且重叠率并不随样本量与技术世代的提升而单调上升。

对这个形势，学界的处理方式高度一致：**它被当作一个数量问题**。样本不够大、人群不够纯、标准不够统一、技术不够灵敏、算法不够强。每一次新技术出

现，同一批人重新出发，理由是「这次不一样了」。七十年过去，这个理由已经用了六七次。

一个持续七十年而形式不变的失败，通常不是因为每一次都做得不够好，而是因为有一条前提从来没有被检查过。本章要检查的就是这一条：

要可靠地改变一个系统，必须先确定这个系统的内部状态。

这句话看上去像常识，像同义反复：不知道它是什么，怎么改它？但在控制理论里，这句话是**错的**，而且它错得可以被精确地证明。**可辨识性**（在给定的输入输出结构下，模型的内部参数能否被唯一确定）与**可控性**（系统能否被输入驱动到任意目标状态）是两个互相独立的结构性质。一个系统可以完全可控而永远不可辨识；也可以完全可辨识而毫不可控。二者之间没有蕴涵关系。

这不是一个哲学立场，是一个可计算的结论。给定一个模型结构与一套观测方案，可以在**收集任何数据之前**判定某个参数是否可辨识——这叫结构可辨识性分析。若一个参数结构上不可辨识，那么无论样本量多大、测量多精确、算法多先进，它都不会收敛；不同的研究会得到不同的估计值，每一个都能通过自己那一份统计检验，而彼此不一致。这正是七十年形势的形状。

于是本章的问题变成这样：**如果「证」在现行的观测结构下结构不可辨识，而辩证论治走的是控制通道，那么七十年的「失败」根本不是失败，而是用错了合格线。**这个说法要成立，需要三件事：需要说清什么在驱动这种不可辨识；需要给出一个别人能在别人的材料上清点出来的读数；需要给出一个能让本章栽跟头的观察。

本章的走法：先摆出三种现成解释并指出它们共同站着的那块地；再从物理学、心理学、工程学各取一条互不相容的单因驱动主张，让三条互相顶住；然后用其中一门自己已经掌握、却从未往这个方向使用的一件事情把三家共同的假定推翻；由此得到承重判断、两根轴与四格、一个可清点的读数、一句不含情态词的判据；最后交出削减、边界、证伪条件、反噬预言与一条写死日期的赌注。

二·把「测不准」这三个字拆开

日常语言里的「测不准」至少混着四件完全不同的事，不拆开就无法讨论。

第一件：精度不足。真值存在且唯一，仪器有噪声，估计值围绕真值散开。处方是提高精度、增加重复。这一类问题会随技术进步而解决，历史上大量案例走的是这条路。

第二件：样本不足。真值存在且唯一，个体差异大，需要更大的样本才能把总体参数估出来。处方是扩大队列。这一类问题会随样本量增加而解决，误差以样本量的平方根收敛。

第三件：人群异质。真值不唯一，因为「这个人群」其实是若干个亚群的混合。处方是分层。这一类问题会随分层的正确而解决。

第四件：结构不可辨识。在给定的观测结构下，存在无穷多组不同的参数值给出**完全相同**的输出。这不是噪声问题，不是样本问题，也不是分层问题——即使数据完美无噪声、样本无限大、人群完全同质，这些参数仍然不能被区分开。

前三件是数量问题，第四件是结构问题。它们的诊断特征完全不同：

- 前三件的表现是**估计值随样本量收敛**；第四件的表现是**估计值随初值与算法而变，与样本量无关**。
- 前三件可以通过「再做一次更大的研究」解决；第四件只能通过**改变观测结构**解决——增加独立的观测通道，或者让输入变得更丰富。
- 前三件的失败会在统计量上留下痕迹（置信区间宽、检验不显著）；**第四件的失败在单项研究内部不留任何痕迹**——优化算法照样收敛到某个点，置信区间照样可以很窄，模型照样拟合得很好。它只在**跨研究比较**时才显现，而跨研究的参数一致性，不是任何一篇论文的报告项。

这最后一条差别是本章全部论述的支点，值得再说一遍：**结构不可辨识在一项研究内部不产生任何异常信号**。你会得到一个漂亮的估计值、一个窄的区间、一个高的拟合优度、一篇可以发表的论文。只有当另一间实验室用同样的方法得到另一个同样漂亮的值时，问题才出现；而那时通行的解释永远是「他们的人群不同」或「他们的方法学质量不同」。

七十年的证候研究，几乎全部资源投在了前三件上。第四件从未被系统检查过。这不是因为工具不存在——结构可辨识性分析在工程学里是成熟的、有软件的、有教科书的；而是因为**没有人把它拿来对准这个对象**，因为没有人怀疑过那条前提。

三·三种现成解释，以及它们共同站着的那块地

在引入三门外学问之前，先把本领域内部最有力的三种解释摆出来，因为本章若不能把它们各自的位置说清，就只是第四种说法。

三之一·复杂性说：层次太多，变量太密

第一种解释说：「证」是一个整体状态，涉及神经、内分泌、免疫、代谢多个层次的耦合，任何单一指标都只能抓到一角；需要的是系统生物学式的多维刻画与足够强的建模能力。

这是七十年里最主流、也最有资源的解释。它的合理性在于对象确实复杂。它的困难在于：**复杂性预测的是慢收敛，不是不收敛**。一个复杂但可辨识的系统，随着观测维度增加，参数估计会越来越稳；而实际形势是维度增加而跨研究一致性没有相应提升。复杂性说没有位置安放这一点，它只能继续要求更多维度。

三之二·异质性说：那不是一个人群

第二种解释说：被判为同一证的病人，其实是一群病因不同的人的混合；在混合人群上找共同指标当然找不到。处方是精细分层——加病种、加分期、加体质、加基因型。

这条路在过去二十年产出很多。它的困难是一个可以观察的形势：**每一次分层都会产生新的、同样不可复现的亚型**。分层没有终点，且每一层的样本量都在缩小，最终落到个案。异质性说无法解释为什么细分之后一致性没有台阶式的改善，它只能说「还不够细」。

三之三·主观性说：判定者不一致

第三种解释把矛头指向识别端：辨证依赖医师的经验判断，一致性有限，以此分组的研究噪声必然大。处方是统一术语、条目化、量表化、四诊仪、辅助判定系统。

这条路的成绩是实打实的：若干病种的判定一致性系数从零点六提升到零点八以上。它的困难有两重。第一，一致性上去了而跨研究的指标一致性没有跟上。第二，也是本章后面要正面处理的一点：**统一判定项这个动作，在结构上恰好会让可辨识性变差**——它削减的正是使参数可分的那种观测多样性。主观性说

的处方与它想达到的目标，在结构上是相反的。这一点在现行框架里完全看不见，因为现行框架里没有「激励丰富度」这个概念。

三之四·三家共同站着的那块地

三种解释分歧很大：一家归给对象的复杂，一家归给人群的混杂，一家归给判定者的不齐。处方也不同：加维度、加分层、加规范。但三家共有一条从不说出的假定：

只要把数据做得足够好，那些量原则上就能被单独确定。

复杂性说假定维度够了就能分开；异质性说假定分层够细就能分开；主观性说假定判定够齐就能分开。三家吵的是**要做到什么程度**，从没有一家问过在**现有的观测结构下**，它们在原理上能不能被分开。

把这条假定念给三家听，三家都会说：这还用说吗？如果一个东西真的在起作用，它当然原则上可以被测出来。

这句「这还用说吗」，正是本章要动的地方。要动它，需要三条互不相容的、各自主张「决定权就在我这里」的强硬立场，把这块地照亮；以及一件出自这三家之一自己的材料，把它顶翻。

四·从三门学问各取一条驱动

选这三门，不是因为它们与中医像。选它们的唯一理由是：三家对「一个系统的内部量由什么决定」给出了互不相容的强硬答案，而且每一家都主张**只有自己那条是决定性的**。凡是说「多种因素共同作用」的立场在这里不算立场，因为它什么也没排除。

四之一·物理学：什么算变量，由标度流决定

问题的形状：一块磁体里有阿伏伽德罗数量级的自由度。要描述它在临界点附近的行为，需要哪些量？答案不是「全部」，也不是由实验者挑选。

物理学在这件事上的强硬主张是：**哪些自由度是相关的、哪些是不相关的，完全由标度变换下的流决定；而这个判定与微观细节无关**。卡丹诺夫的区块自旋与威尔逊的重整化群把这条说死了：在标度变换下，某些耦合的系数随粗粒化而增长（相关算符），某些衰减到零（不相关算符）。衰减的那些，无论初始值多

大，在长波长极限下都不影响宏观行为。由此得到普适类——液气临界点与三维伊辛磁体的微观机制毫无共同之处，而它们的临界指数在实验误差内完全相同。

这条主张的强硬之处在于它取消了「先有真实变量清单，再去发现它」这个图像：变量清单是被标度流选出来的，而选出来的结果依赖于你在什么尺度上看。

它的动力式是：

宏观上能被观测到的显露，与观测所处的尺度和条件之间的矛盾，逼出「哪些量算相关变量」这条组织路径改变；新的相关变量清单回过头改写了该在什么尺度上观测、什么算一次显露。

时序读数：先动的通常是条件一侧——维数变了、对称性变了、加入了长程相互作用；相关变量清单在理论上很快跟上；跟不上的永远是实验探针，因为新的序参量造一台仪器要十年。

这一家自曝的一处：被判为不相关的算符，在有限尺寸的实际系统里确实产生可测量的修正——所谓修正指数。也就是说，「不相关」是一个取极限之后的结论，而任何真实系统都不在极限上。物理学知道这件事，把它写在修正项里，从不把它读作「被丢掉的那一份仍在起作用」的证据。

四之二·心理学：构念的边界由参照条件决定

问题的形状：一个人「焦虑水平高」。高，是相对于谁？相对于哪个常模、哪个年龄段、哪个文化、哪套题目？

心理学有一支的强硬主张是：一个心理构念是否成立、边界划在哪里，完全由它所处的参照条件场决定——常模群体、题目集合、施测情境、诊断类别的划分惯例；换一个参照条件，构念本身就不是原来那个了。测量不变性的研究把这件事变成了可检验的：同一份量表在两个群体上是否测量同一个东西，可以用一组嵌套模型逐层检验，而大量实际量表通不过其中较严的层级。若通不过，那么两个群体上的分数不可比较——不是「差异小」，是没有共同的尺子。

它的动力式是：

被试当下显露出的表现，与施测与临床所用的操作路径之间的矛盾，逼出这个构念赖以成立的条件场改变；新的条件场回过头改写了什么表现能被记录、以及该用什么操作去测。

时序读数：先动的通常是操作路径一侧——出现了新的施测方式（在线施测、生态瞬时评估、被动传感）；构念的参照条件在一到两个修订周期内跟上；跟不上的是教科书与培训体系，常常滞后十年以上。

这一家自曝的一处：聚合原则。单次行为在跨情境上的一致性系数长期徘徊在零点三左右，曾经被读作「特质不存在」；而把足够多次行为聚合起来之后，一致性回升到零点七以上。这说明存在某种不随参照条件漂移的成分，只是它在单次观测里不显露。心理学知道这件事，把它用来为特质研究辩护，从不把它读作「某些量只能以组合的形式存在」的证据。

四之三·工程学：状态显露成什么样，由激励与观测结构决定

问题的形状：一个化工反应器，内部有若干看不见的状态。装了三个传感器，做了实验，拟合出一组参数。这组参数是真的吗？

工程学的强硬主张是：**一个系统的内部状态显露成什么样，完全由输入的激励设计与观测结构决定；若输入不够丰富——不满足持续激励条件——那么参数估计不会收敛到任何确定的值，而「系统的真实参数」这句话在此时没有操作意义。**这不是经验之谈，是可以在采集数据之前算出来的：给定模型结构与观测方案，结构可辨识性分析会告诉你哪些参数可以被唯一确定、哪些只能以组合的形式被确定、哪些完全不能。

这条主张最狠的地方是它把责任从数据转移到了**结构**：不可辨识不是数据不够的表现，它是先验的。

它的动力式是：

辨识路径（输入设计、采样方案、模型结构）与系统实际运行的条件场之间的矛盾，逼出这个系统显露成什么状态；新的显露回过头改写了该怎么设计激励、以及系统被允许在什么工况下运行。

时序读数：先动的通常是运行条件一侧——工况迁移、原料变了、负荷变了；可观测的显露随即改变；跟不上的是模型结构，因为改模型要重新验证、重新认证。

这一家自曝的一处：这一处是本章的枢纽，下一节整节处理。

四之四·三家不能同真在哪一点上

三家各自主张一条驱动，而被驱动的那一端互不相同：物理学说被驱动的是**变量清单**，心理学说被驱动的是**构念赖以成立的条件场**，工程学说被驱动的是**系统显露出的状态**。

这不是三个角度，是三条不能同真的主张。若物理学对，那么在工程学的场景里，哪些参数算数应当由尺度分离决定，而不由激励设计决定——激励再丰富也测出不相关的量。若工程学对，那么在物理学的场景里，相关变量的判定应当依赖于你能施加什么样的扰动，而不依赖于标度流——而这正是物理学明确否认的。若心理学对，那么在另两个场景里，决定权应当在参照群体与惯例手里，而这在物理学看来荒唐。

三条不能靠判谁对化解，因为三家不在同一个语域里作答。唯一的出路是承认一件三家都没说的事。要说这件事，需要材料，而材料必须来自三家之一自己。

五·翻转发生在哪一处：工程学自己的那件材料

材料在工程学手里，而且是自适应控制这门学问最基本的一段结果，任何研究生课程都会讲。

自适应控制要解决的问题是：被控对象的参数未知，控制器一边估计参数一边控制。直觉上，这件事的成败应当取决于参数估计得准不准——估准了才能控制好。整个领域早期的努力也是朝这个方向去的。

结果不是这样。经典的稳定性证明给出的结论是：**在合适的自适应律下，闭环跟踪误差趋于零，而参数估计并不趋于真值**。参数收敛需要一个额外的、独立的条件——输入必须满足持续激励。若输入不满足，参数估计会沿着一个流形漂移，永不收敛到某个点；而与此同时，输出跟踪照样收敛，控制目标照样达成。

也就是说：**在这门学问自己每天使用的核心结果里，控制的成功与辨识的成功是脱钩的**。

更要紧的是随之而来的标准做法。当结构可辨识性分析告诉你某些参数不可单独确定时，工程师的做法不是停工，是**重参数化**：把不可辨识的原始参数重新组合成一组可辨识的组合，用这组组合改写模型，然后照常设计控制器、照常预测输出、照常通过验证。

请注意这一步做了什么。原始参数里的某几个，从此不再作为单独的量存在于模型中。它们没有被测出来，也没有被判定为零或无关；它们被吸收进了一个组合里。模型的自由度变少了，一切都干净了，验证通过了，论文发表了。而那几个量在物理上依然存在、依然在系统里各自起作用——只是从这一刻起，账本上没有它们的名字了。

这一步在工程学自己的领域里回答的是另一个问题：如何得到一个良态的、可估的模型。方向翻没翻，与那个问题无关，所以没有人往那儿看。

这件材料一旦被摆到这个位置上，三家的共同假定就顶不住了，而且顶不住的原因不是有人反对，是反例出自三家中最讲究可测量性的那一家自己：那家主张「激励与观测结构决定状态显露」的工程学，自己的核心结果里存放着一整套不需要状态被确定就能可靠改变状态的方法。

于是可以说出那句被共有前提挡住的话：

辨识与控制是两条互不隶属的通道。方向不是固定的：这一轮可以由控制先动，辨识永远不来，而辨识的不来不产生任何控制上的后果；下一轮也可以由观测结构先动，逼着控制律重写。

三家看见的是这个循环里的不同段落：物理学看见的是「条件先动、变量清单跟上」那一段，心理学看见的是「操作先动、参照条件跟上」那一段，工程学看见的是「工况先动、显露跟上」那一段。三家都没错，三家都只看见了自己那一轮。

六·承重判断：消参

现在可以把判断写出来。

「证」不是一个尚未被找到序参量的复杂状态，不是一个尚未被细分干净的异质人群，也不是一个尚未被统一的主观判定，而是一类在现行观测结构下结构不可辨识、而在控制上完全够用的量。它在辨识账本上被并入某个可辨识组合、因而不作为单独的东西存在；它在控制账本上每天单独起作用，因而从不缺席。

被并进去的那一份，本章称为消参。

这个名字要担的意思有两层，缺一层就不成立。

第一层：它不是被证明为零，也不是被判为无关，它是被吸收。这三者在账本上的处理完全不同。被证明为零的量，账上有它，值是零。被判为无关的量，账上有它，标着不相关。**被吸收的量，账上没有它。**模型里剩下的是那个组合，组合有名字、有估计值、有置信区间；构成它的那几个原始量，从此在这套形式体系里不构成一个可指称的对象。

第二层：它在控制上单独起作用。这一层把消参与真正的冗余参数分开。一个真正冗余的参数，改变它不会改变系统的任何行为，删掉它没有代价。消参不是这样：改变构成组合的某一个量、同时反向改变另一个量，使组合保持不变，输出确实不变——**但系统对扰动的响应、它的稳定裕度、它离失稳有多远，全都变了。**也就是说，在稳态的输入输出层面它是不可见的，在扰动与工况迁移时它是可见的。这一点后面会被写成一条可裁决的预测。

有了这两层，开头那件怪事就有了单一的来源，不必再拆成两半解释。

它不可还原，是因为结构不可辨识与数据量无关。七十年里每一轮研究都在做同一件事：在一个固定的观测结构下（望闻问切四类通道加若干实验室指标）估计一组内部量。若这些量在该结构下只能以组合的形式确定，那么每一轮都会得到一个自己的解——每一个解都能拟合自己的数据，每一个解都不同。样本量增加只会让每个研究的置信区间变窄，不会让不同研究的解靠拢。**这正是七十年**

的形势：区间越来越窄，彼此越来越不重合。

它有效，是因为控制不需要辨识。一位医师并不需要知道病人体内每一个量各是多少，他需要知道的是：施加这一组动作之后，输出往哪个方向走。辩证论治积累的正是这种输入输出映射——数千年、数亿诊次的闭环调整，是一套极其庞大的经验控制律。它从未走过辨识通道，所以它交不出参数；它一直在走控制通道，所以它交得出结果。

为什么这三家自己到不了这一步。不是疏忽。物理学的核心问题是普适行为，它一旦承认「被判为不相关的那一份在实际系统里单独起作用」，就必须放弃标度分离带来的全部简化，而那是它最大的资产；停在修正指数上，对它是最经济的。心理学的核心问题是分数的可比较性，它一旦承认存在只能以组合形式存在的构念，就无法向使用者交代分数的意义；停在测量不变性的检验上，对它是最省事的。工程学的核心问题是把系统控制住，它一旦把「被吸收的参数」重新立为对象，就得为它承担建模与验证的成本，而这些成本换不来任何控制性能

的提升；它有充分的理由不去立。一个领域看不见什么，通常不是它的缺陷，是它的问题结构的影子。

七·两根轴与四格

一个名字不能分东西，两根轴才能。

- 轴一：在现行观测结构下，内部量能否被唯一确定。两端是「可辨识」与「不可辨识」。这是一个可以在收集数据之前算出来的结构性质。
- 轴二：这个系统能否被可靠地驱动到目标状态。两端是「不可控」与「可控」。

	不可控	可控
可辨识	② 知而不能改：年龄、基因型、既往暴露、大多数预后标志物。测得准，改不动。这一格是流行病学与预后研究的主场。	① 还原医学的主战场：血糖、血压、病原体载量、电解质。测得准，也改得动，测与改共用同一个量。
不可辨识	③ 真噪声：既分不开也用不上 的量。这一格里的东西应当被 丢弃，丢弃没有代价。	④ 消参：「证」、心理治疗中 被称为共同因素的那些量、复 杂装置上老师傅的调参手感、 大模型里的内部表示。改得 动，说不清；说不清不影响改 得动。

四格必须落到具体的人和事，否则这张表只是修辞。

第①格是七十年里被当作唯一合格标准的那一格，也是还原医学真正成功的地方。它有一个可核查的签名：跨研究的参数估计随样本量而收敛。

第②格常被误当作第①格的成功案例，但它在临床上是另一回事。一个测得极准的预后指标，若没有任何可及的手段改变它，它的全部价值是分层与告知。把资源大量投在第②格而以为在做第①格的事，是当代精准医学一个未被充分讨论的实际问题。

第③格是应当被丢弃的。要紧的是：第③格与第④格在任何单项辨识研究内部看起来一模一样——两者都表现为参数估计不稳。区分它们的唯一方法是去看控制端：这个量参与的处置有没有可复现的效果。而辨识研究从不看控制端。

第④格是本章要打的那一格。它与第③格只差一个可控性，与第①格只差一个可辨识性，而现行的评价体系只有一根轴——它把第④格与第③格一起判为失败。

八·为什么第④格能通过全部形式审查

一个值得打的靶，必须是所有现成检查都发现不了的那一格。第④格满足三条签名。

签名一：短期指标全面改善。

第④格里的每一项常规指标都可以很好看，而且是真的好看。控制端：疗效可以重复，满意度可以高，临床路径可以顺畅。辨识端：**每一项单独的研究都能产出阳性结果**——不可辨识不妨碍优化算法收敛到某一点，也不妨碍那一点通过统计检验。于是文献数量上升、影响因子上升、被引上升。规范化端：判定一致性上升，术语统一，可教可考。

在任何一项审查里，第④格与第①格看起来一模一样。

签名二：失效不产生任何信号。

这一条是第④格最难对付的地方。设想一次真正的失效：某个在控制上起作用的量被吸收进组合之后，后续二十年里所有的模型都不再包含它；某一次工况迁移（新的疾病谱、新的合并用药、新的生活方式）使这个量的取值移出了原来的范围，于是那个组合虽然还是同一个值，系统的实际行为却变了。

这次失效会在哪里被记录？不会在任何地方。不会在模型验证里——组合值没变，拟合照常。不会在临床记录里——每个病人都被判了证、开了方、记了转归。不会在文献里——阴性结果被归因于样本或人群。不会在标准修订说明里——修订记的是条目增删，不记「本次修订改变了哪些量的可辨识性」。

跨研究的参数不一致，不出现在任何一份报告的任何一个字段里。这不是有人隐瞒，是没有这个字段——因为报告规范要求的是本研究内部的精度，从不要与既往研究的参数一致性。

签名三：自我加固，越正规化越严重。

这是本章最反直觉、也最容易被检验的一条。

结构可辨识性依赖于观测的**多样性**：不同的量必须能在数据里独立地变化，才可能被分开。若两个观测项在所有记录里总是同时出现、同时缺失，那么它们对应的参数在数学上就不可能被分开——无论有多少条记录。

而规范化做的是**什么**？统一问诊提纲、统一四诊项目、统一记录模板、统一判定流程。这些动作全部指向同一个方向：**让每一次观测都尽可能相同**。相同正是它的目的，因为相同才有一致性。

但相同同时意味着**激励贫乏**。一套高度标准化的问诊，其观测项之间的共现结构是被模板固定住的；模板里绑在一起的项，在数据里永远绑在一起。于是：

判定一致性上升的过程，同时是结构可辨识性下降的过程。

规范化程度与所有常规质量指标正相关，与本章关心的那个性质负相关。**做得越好，可辨识性越差，而所有的表都显示做得越来越好。**

这一条不是推测，它可以在任何一家医院的病历数据库里被算出来：统计入判项的两两共现结构，看有多少对是完全共线的，再比较规范化程度不同的机构。数据已经存在，不需要新的采集。

九·一个可以清点的读数：消参率

一个判断如果只能被同意或反对，它就还没落地。

九之一·定义与编码规则

消参率 = 在一个判定模型（或一套判定规则）中，原始参数的数目与在给定观测方案下可被唯一确定的独立参数组合的数目之差，占原始参数数目的比例。

它是一个比例，无量纲，跨领域可以并排比较。它有一个别的读数少有的优点：**它是纯结构计算，不需要任何数据**。给定模型结构与观测方案，在做第一次实验之前就能算出来。这使它成为本章全部检验里最便宜的一个。

清点它需要三条编码规则，缺一条这个数不可复算：

- **规则一：什么算一个原始参数。** 在判定规则里被赋予独立含义、并在文本中被单独命名的量。一条判定规则若写着「舌淡、脉沉细、畏寒，三者合并权重为二」，那么这里三个原始参数还是一个，取决于规则文本是否允许它们各自取不同的值。规则文本说了算，不由分析者事后合并。

- **规则二：什么算「可被唯一确定」。**在给定的观测方案下，若存在两组不同的参数取值给出完全相同的全部可观测输出，则相应参数不可唯一确定。判定用结构可辨识性分析的标准方法，与数据无关。
- **规则三：什么算一个观测方案。**记录模板中允许独立取值的观测通道清单，加上这些通道在实际记录中的共现约束。**共现约束必须从真实记录中统计，不能从模板的理论可能性中推定**——模板允许而实际从不独立出现的通道，在结构上等同于不存在。

第三条是三条里最容易被绕过、也最要紧的一条。一份问诊模板在纸面上有六十个可填项，而实际记录中只有十二种填法反复出现，那么它的有效观测通道是十二个量级的事，不是六十个。

九之二·四个领域里的同一个比率

药代动力学：一个二室模型有四个参数，若只测中央室浓度，其中某些参数只能以组合形式确定。教科书里的标准例子给出的消参率在零点二五到零点五之间，**这个数在采血之前就算得出来。**

心理测量：一个含有若干潜变量与残差相关的结构方程模型，其识别状态可用标准规则判定；未被识别的自由参数数占总自由参数数的比例即是它的消参率。大量已发表模型是通过施加「便利性约束」（把某个载荷固定为一、把某个残差相关设为零）达到识别的——**这些约束本身就是一次消参操作，而它们通常写在方法部分的一句话里。**

工业过程建模：一个机理模型有数十个动力学与传热参数，而在正常工况的数据上，其中很大一部分只能以组合确定，因为操作规程把若干操纵变量绑在一起变化。这个数在工程实践中是已知的，工程师用重参数化处理它，不把它当作一个需要报告的量。

中医证候判定：取一版证候诊断标准的条目与权重，写成一个判定模型；取该院实际病历中入判项的共现结构作为观测方案；算消参率。据本章所知，**这件事从未被做过一次。**它不需要新的病人、新的样本、新的伦理审批，只需要一版标准文本和一批脱敏病历。

四处的量纲完全不同——房室参数、载荷、动力学常数、判定条目——而比率的定义是同一个。

九之三·与既有主要指标正交的那个例子

一个新读数若与既有主要指标高度相关，会被立刻读成那个指标的代理。所以必须举出一个「既有指标最好看、而这个数最难看」的实例，举不出来就该重造。

例子在系统辨识的教科书里：一个模型对数据的拟合优度达到零点九九，残差检验全部通过，交叉验证预测精准，被工业界长期采用——而它的消参率可以是零点六。六成参数无法单独确定，只能以组合确定。**拟合优度与可辨识性之间没有任何蕴涵关系：**不可辨识意味着有无穷多组参数给出同样好的拟合，因此不可辨识的模型恰恰更容易拟合得好，因为可行解的集合更大。

同样的关系出现在证候研究里：一版判定一致性系数最高的标准，其消参率也最高，因为一致性是靠把观测项绑成固定组合换来的，而绑在一起的项永远分不开。两个数在方向上相反，任何现行的质量框架都不会同时报告它们。

九之四·这个数的伦理边界

消参率会被拿去评价模型，进而评价做模型的人，所以三条必须写死：

其一，这个读数指的是结构，不是人。高消参率描述的是「这套模型与这套观测方案之间的关系」，不描述任何研究者的能力。事实上，一个诚实地报告了高消参率的研究者，比一个用便利性约束把它掩盖过去的研究者做得更好。

其二，不得为了把这个数做好看而删参数。通过删除不可辨识的参数来把消参率降到零，是本章最担心的误用，见第十七节的第一条反噬。任何以降低消参率为目标的评审规则，都必须同时要求报告删参前后的控制性能。

其三，已经据此作出不利评价的，必须公开三条编码规则并给出复核通道。消参率的大小完全取决于规则三对「有效观测通道」的本地化认定，任何据此作出的评价都必须允许被评价者按公开规则复算。

十·一句不含情态词的判据

一个判断要能被用，得配一句可以拿去问记录、而不是拿去问人的判断的话。

本章的判据是：

在这一次判定所依据的全部观测项里，有多少对在过去 N 次记录中始终同时出现、同时缺失？

这句话里没有「应当」「充分」「真正」「合理」。它问的是共现结构，答案是一个数，由数据库直接给出。它之所以是本章的判据，是因为完全共线的观测对，是结构不可辨识最直接、最容易清点的来源。

把它拿到五个场景里跑一遍，五个答案互不相同：

1. **一家中医院的门诊病历库**：统计现行判定标准所列入判项在近三年记录中的两两共现，数出完全共线的对数。答案是一个整数，且本章预测它在规范化程度更高的科室里更大。
2. **一个药代动力学采样方案**：给定采样时点集合，算出哪些参数只能以组合确定。这个答案与病历无关，是纯粹的结构计算，且在同一个模型下随采样时点的增删而跳变。
3. **一份心理量表的施测数据**：统计题目间在样本中局部依赖程度达到完全共变的题对数。这个答案通常不大，因为量表编制本身会主动删掉共变过高的题——这正是心理测量比中医判定多做了一步的地方。
4. **一家工厂的过程历史数据**：正常工况下从未独立变化过的操纵变量对数。这个答案往往很大，因为操作规程明确要求若干变量联动。工程师知道这个数很大，用重参数化处理，不报告。
5. **一套课堂评价的评分维度**：评分者从未给出不同分数的维度对数。这个答案在使用整体印象评分的场合接近全部维度对，在使用分维度独立评分且有防串扰设计的场合显著下降。

五个答案里，至少有两个不能从本章的判断直接推出来：心理量表在编制阶段主动删共变题这件事、以及工业过程里被规程绑死的变量对数的实际量级——这两条只能查，不能想。若五个答案全都能从命题推出，那它就不是判据，只是命题的复述。

十一·在别的领域里能预测出什么

类比是「像」，兑现是「在那个领域里也能据此预测出一件具体的事」。以下十二处，每一处都给出一个可以去查的预测。

一、**系统辨识**：对同一个机理模型，在采样方案扩容（增加独立可变的激励通道）之后，跨实验室的参数估计离散度应当下降；而在样本量扩大而激励结构不变时，离散度不下降，只有各自的置信区间变窄。这是本章全部主张里最直接的一条，可以在既有数据上复核。

二、**药理学**：同一药物的群体药代模型，其跨研究参数差异应当与各研究采样方案的差异强相关，而与样本量弱相关。这个检验可以在已发表的群体药代文献上做一次元回归。

三、**自适应控制的实践**：在持续激励条件不满足的现场系统中，控制性能可以长期良好，而参数估计漂移；本章预测这类系统在**工况迁移**时的性能跌落幅度，显著大于参数已收敛的同类系统。这是第十二之三节那处削减的直接来源，也是本章对中医急症弱势的解释。

四、**心理治疗研究**：若各流派的疗效差异确实来自「多组等价解」而非「共享的共同因素」，则**严格控制住共同因素之后，流派间差异仍应为零**。共同因素论预测差异此时应当显现。这是一个可以设计的实验。

五、**心理测量**：通过测量不变性检验的构念，其消参率应当显著低于未通过的构念；而一个构念的效应量大小与它的消参率之间应无相关。前者是本章的预测，后者是本章的零假设，两条一起构成一个可判的组合。

六、**可解释机器学习**：对同一个高性能模型，不同的归因方法给出的特征重要性排序应当系统性地互相矛盾，且矛盾程度随模型的消参率上升而上升。这一条已有大量实证支持，本章提供的是解释：**归因方法的不一致不是方法缺陷，是被解释对象在该观测结构下不可辨识的表现**。

七、**统计物理**：人为改动被判为不相关的耦合，宏观临界指数不变（物理学的预测）；但本章预测**系统对外部扰动的响应函数会改变**，且改变幅度随系统尺寸减小而增大。这是一个可以在数值模拟里做的检验，也是本章与不相关性判定之间的分离线。

八、**生态系统管理**：一个食物网模型的参数在观测数据上大量不可辨识，而基于它的管理干预可以有效。本章预测：这类干预在**环境条件迁移**时的失败率，显著高于建立在可辨识参数上的干预。

九、**教育评价**：使用整体印象评分的体系，其维度间共线对数高、消参率高，而评分一致性也高；改用带防串扰设计的分维度评分后，一致性下降而消参率下降。本章预测这两个数在改革前后向相反方向移动。

十、大语言模型的内部表示：模型内部的表示单元在给定探针集合下大量不可唯一确定，而基于表示的干预（激活引导）可以有效。本章预测：干预的有效性与探针能否唯一确定表示之间没有相关，而干预在分布迁移下的稳健性与之相关。

十一、临床指南：若一条推荐所依据的模型消参率高，则该推荐在不同人群中的实际效果方向应当更容易反转。本章预测指南推荐的「跨人群一致性」与其底层模型的消参率负相关。

十二、中医经典方剂：一张多味方的各味药的独立贡献，在既有的整方给药观测结构下结构不可辨识——这是拆方研究七十年产出稀薄的结构性原因，而不是拆方研究做得不好。本章预测：只有引入**能让各味药独立变化的给药方案**（析因设计而非整方对照）才能改变这一点，而增加样本量不能。

十二·三处反过来给本判断加的约束

本章的判断不是从三门学问里取长补短得来的，所以也必须接受三门学问反过来的削减。以下三处，每一处都真的削掉了本章原本打算主张的东西。

十二之一·物理学削掉了「不可辨即只能靠试错」

重整化群给出的不只是「微观细节不相关」这个否定结论，还有一个极强的肯定结论：**同一普适类的临界指数可以被精确计算，并与实验在小数点后若干位上相符**。也就是说，在无法确定任何微观参数的情况下，粗粒化层可以有严格的、定量的、可预言的理论。

如果没有这一处，本章原本会写下的更强的一句是：「结构不可辨识的系统只能靠经验试错，理论化的努力应当让位于疗效验证。」这句话现在不能写了。物理学证明了相反的可能：**不可辨识的对象在合适的粗粒化层上完全可以被严格定量化**。

被削掉的因此是本章的一个实践主张。本章不能主张停止对证的理论化，只能主张**理论化应当在组合层而不是在原始参数层上进行**——去建立可辨识组合之间的定量关系，而不是继续去分解组合。这比原来的主张弱，也更有建设性。

十二之二·心理学削掉了「追求跨研究参数一致是浪费」

测量不变性研究给出了一套可检验的、分层的判据：形态不变、载荷不变、截距不变、残差不变。有些构念确实通过了较严的层级，而通过的那些，其跨群体分数比较是有依据的。也就是说，**可比性不是全有或全无，它是分层的、可检验的、部分可达成的。**

如果没有这一处，本章原本会写下的更强的一句是：「既然结构不可辨识，追求跨研究的参数一致性就是把力气花在一个不可能的目标上。」不能写了，而且被削掉的不只是一句主张，是本章的价值排序。

被迫承认的是：**消参不是一个二分，是一个连续谱**，而消参率正是这个谱上的位置读数。一套判定规则里可以有一部分参数可辨识、另一部分被吸收；正确的做法是把两部分**分开报告**，而不是笼统地宣布这个对象不可还原。第④格因此不是一个抽屉，是这个谱的一端。

十二之三·工程学自己削掉了「可控就够了」

自适应控制的完整图景比第五节引用的那一段更复杂。参数不收敛时，闭环虽然稳定，但会出现若干已知的病态：在小扰动持续存在时估计值可能缓慢漂移直至突然失稳（这一现象在文献里有专门的名字），瞬态性能变差，对未建模动态的鲁棒性下降。工程实践为此发展了一整套修补手段——投影、死区、泄漏项、正则化——每一项都是在为「参数没收敛」付账。

如果没有这一处，本章原本会写下的更强的一句是：「辨识对控制无关紧要。」不能写了。正确的表述是：

不可辨识的控制在稳态下够用，在扰动与工况迁移时会显著失效，且失效方式常常是突然的而不是渐进的。

这一处削减不但削弱了本章，还给了本章一个它原本没有的解释力：它恰好对应着辨证论治在**慢性、稳态、长期调理**场景中的相对优势，与在**急症、大扰动、快速恶化**场景中的相对弱势。这个强弱格局在临床上是公认的，此前一直归因于手段的快慢；本章给出的是一个结构性的来源。**这条削减因此同时是本章最有力的一处兑现。**

十三·与最近的几种说法的分界

本章的判断有一个明显的危险：它听上去像好几种已有的说法。逐一处理，每一位都写出「看到什么就说明本章对而它错，看到什么就说明反之」。

结构可辨识性分析本身 (Bellman & Åström, 1970; Walter & Pronzato, 1997; Villaverde 等的近期工作，工程学与系统生物学)。这是本章最危险的一位，因为它不是内容上的近邻，是**方法上的**——它已经有一整套完整的判定理论、软件和几十年文献。若本章只是把它搬到中医里，那本章就是被它覆盖的。分离线在**目的**：可辨识性分析把不可辨识当作**待修复的实验设计缺陷**，其标准处方是改进激励、增加观测、或重参数化后重报；本章主张存在一类对象，其不可辨识不必修复、且修复的代价可能是删掉在控制上起作用的量。判定形式：若在**扩容观测使某模型恢复可辨识之后，其控制性能显著提升，则可辨识性分析的处方对**；若恢复可辨识的唯一可行途径是删参数、且删参后控制性能下降，则本章**对**。这一条可以在工业过程模型上直接做。

持续激励条件与自适应控制 (Åström & Wittenmark, 1995; Ljung, 1999, 控制论)。第二位方法学占位者。它已经说清了「激励不足则参数不收敛而控制照常」。分离线在于**激励不足的来源**：控制论把它当作实验设计的可选项——你想让参数收敛，就去设计更丰富的输入；本章指出在临床里，激励贫乏不是可选项，而是**规范化制度的直接产物**，因而它不是可修的设计缺陷，是结构性的、且与所有质量指标同向。判定形式：若**规范化程度更高的机构，其入判项两两共现率更高，则本章对**；若两者无关或反向，则激励贫乏只是设计疏忽，控制论的处方够用。数据已存在。

重整化群与普适类 (Kadanoff, 1966; Wilson, 1971; Fisher, 1974, 物理学)。分离线在于被丢掉的那一份有没有作用：普适类主张不相关算符对宏观行为**没有影响**（在极限意义上），本章主张被吸收的那一份在**控制与扰动响应上有影响**，只是在稳态输入输出上不可见。判定形式：若人为**改动被判为不相关的量，而系统的扰动响应函数随之改变，则本章对而「无影响」的读法过强**；若响应函数完全不变，则本章错。数值模拟可判。

多重可实现性与特殊科学的自主性 (Putnam, 1967; Fodor, 1974, 哲学)。这是听上去最像的一位：心理种类由功能角色定义，同一功能可由多种物理基质实现，因此心理学不可还原为神经科学。分离线在于**多在哪里**：多重可实现说的是一对多的**实现关系**（同一功能，多种基质），跨基质、跨物种；本章说的是一对多的**辨识关系**（同一输出，多组参数），可以完全发生在**同一个基质、**

同一个个体上，而且可以算出等价类的维数。判定形式：若在单一个体、单一基质上仍出现参数不可唯一确定，则本章说的是另一件事；若不可确定性只在跨基质比较时出现，则本章不过是多重可实现的一个说法。药代模型的例子直接落在前者。

共同因素论（Rosenzweig, 1936; Wampold, 2001, 心理治疗研究）。本领域之外最近的一位：不同流派疗效相当，特异性技术的贡献很小，因此起作用的是各流派共有的成分——治疗同盟、期待、仪式。它与本章对同一批证据给出不同的读法。分离线：共同因素论认为**存在一个共有成分**，各流派都含有它；本章认为**不存在共有成分**，各流派是同一个不可辨识系统的不同等价解。判定形式：若在实验上严格控制住所谓共同因素之后，流派间的效果差异仍然为零，则本章对——因为共同因素论预测此时差异应当显现；若差异显现，则共同因素论对。这条是两说之间最锋利的一刀，且可以设计实验。

因果效应的可识别性（Pearl, 1995; 因果推断文献，统计学）。分离线在于**纲领的方向**：可识别性理论的整个工作是寻找足以使效应可识别的假设与设计，其隐含承诺是「找到合适的假设就能识别」；本章主张存在一类系统，其内部量在任何可检验的假设集下都不可识别，而干预照样可靠。判定形式：若**存在一个干预效果稳定可重复、而其内部量的效应在任何可被数据检验的假设下都不可识别的案例**，则本章对；若每一个这样的案例最终都被证明只是假设选得不好，则**可识别性纲领够用**。

默会知识（Polanyi, 1958, 1966, 认识论）。第④格里的典型例子——老师傅的手感——正是默会知识的标准案例，所以必须正面处理。分离线在于**说不出来的原因**：默会说归因于不可言说，即那份知识在原理上抵抗形式化；本章归因于可言说而不可唯一确定——被吸收的那一份完全可以被写下来，写成一组参数组合与约束，它只是不能被拆成单个参数的值。判定形式：若把一位老师傅的调参经验完整地形式化为一组可执行的约束（说得出来），而仍然无法从这组约束反解出各个参数的单独取值，则本章对而默会说错；若形式化本身就做不到，则默会说对。这条在工业现场可做。

不可通约（Kuhn, 1962, 科学史）。剥掉全部医学名词后，本章的形式是「一套评价体系用它自己的合格判据去评价一个不按该判据运行的对象，于是该对象被长期判为失败而实际一直在成功」。这个纯形式的结构在科学史里最近的名字是不可通约。必须点名，因为读者极易把本章读成它的一个应用。分离线

在于是否需要诉诸语言：不可通约主张两套范式的术语无法互译、无中立语言可作裁决；本章主张辨识与控制两个性质可以在同一套数学里被精确定义，并被证明互相独立——不需要任何不可通约，只需要一条被忽略的定理。判定形式：若二者的独立性能在同一形式体系内被证明并计算，则不必诉诸不可通约，本章对；若这种独立性只能在语言层面被谈论而不能被计算，则本章其实是不可通约的一个变体。这一条本章当场兑现：消参率是可计算的。

可解释性研究中的「事后解释不可靠」（Rudin, 2019，机器学习）。它说：对黑箱模型的事后解释常常不忠实，应当改用本身可解释的模型。分离线：它把不忠实归因于解释方法的缺陷，处方是换模型；本章归因于被解释对象在该探针结构下不可辨识，处方是改探针结构或接受组合层解释。判定形式：若不同归因方法的不一致程度随模型消参率上升而上升，则本章对；若不一致程度与消参率无关而只与解释方法的技术优劣有关，则可解释性研究的诊断对。

证候客观化纲领（沈自尹自一九五九年起的系列研究、陈可冀等的血瘀证研究、证素辨证一路，中医学）。分离线在于对收敛的预测：客观化纲领预测跨研究一致性随样本量与技术世代而上升；本章预测一致性的上限由消参率决定，与样本量无关。判定形式：若跨研究参数离散度随样本量下降而与消参率无关，则本章错；若离散度与消参率强相关而与样本量弱相关，则本章对。这条检验所需的全部数据在既有文献里已经存在。

最近的那一位是谁。是结构可辨识性分析本身。它比本章早半个世纪，工具更完备，而且它已经说过「不可辨识与数据量无关」这句话。本章之所以仍不能被它吸收，只有一个理由：它把不可辨识一律当作待修复的缺陷，因而它的整个工具箱里没有一个位置安放「不必修复的不可辨识」。一旦承认这种东西存在，它的标准处方——改进设计直到可辨识——就不再是普遍适用的了。若将来有人证明所有临床上有效的不可辨识都可以在不损失控制性能的前提下被修复，那么本章应当被撤回，可辨识性分析足够用了。

十四·与本系列各章的分界

本系列近期几篇处理相邻的题目，其中三篇用的是同样这三门学问，一篇处理的是同一道题。它们是本章最近的邻居，必须逐篇说清并各给一条能分开的预测。

《随程》（第七章）（同一道题，从教育学、化学、艺术学入手）。那一篇的判断是：证是一类只在识别程序运行期间成立、并会回写该程序的对象，七十年不收敛是因为每一轮研究在结束前已把自己的对象重划过一次，读数是随程率。本章的判断是：证在给定观测结构下结构不可辨识，而不可辨识与可控互相独立，读数是消参率。**两篇的位置不同**：那一篇处理的是**对象的边界随时间被重划**，本章处理的是在任一时刻的**固定边界内，内部量本来就分不开**。

判决性对照：取一个诊断标准数十年未修订的封闭体系——随程率为零。

《随程》预测该体系内的机制不发动，跨研究指标应当收敛；**本章预测仍不收敛**，因为消参率由观测结构的共现约束决定，与标准是否修订无关。这一例能把两篇彻底分开，而且可查：找一个长期未修订的证候标准或一个稳定的经典方证体系，看其标志物研究是否收敛。

两篇的关系是互补而非重复：随程率高导致**跨轮次**不可比，消参率高导致**轮次内部**不可辨。二者叠加时，跨研究离散度不是相加而是相乘——因为每一轮内部先散开一次，轮与轮之间再错位一次。**这也意味着两篇给出的处方不冲突但顺序不能颠倒**：先降消参率（扩容观测通道）再降随程率（记录归类变更），反过来做没有意义，因为在不可辨识的结构里记录再多的归类变更也拼不出参数。

《阈前距》（《治未病立账学》第二章）（何谓未病，同三学科）。那一篇说的是：判定线本身在动，一个人与线之间的距离变化中有多少来自线，账本不设字段，读数是线移占比。**分离线**：那一篇的对象是**判定阈值**，本章的对象是**内部参数**。**判决性对照**：一条二十年未动的判定线（线移占比为零）之下，《阈前距》的机制不发动；本章预测该体系内的参数仍不可唯一确定，跨研究估计仍不一致。两篇在这一例上给出不同结论。

《悬量》（《治未病立账学》第八章）（治未病为何实践最少，同三学科，同为三原理型）。那一篇说的是：阈下风险总量确定而归属未定，一切介入必须落在具体对象上，故只能落在构造出来的对象上，读数是落点自证率。**分离线**：那一篇处理的是**下不了手**（没有可落点的对象），本章处理的是**下了手也说不清是哪一下起的作用**（对象在手而参数不可分）。**判决性对照**：一个症状已显、落点明确的病人（落点自证率高），《悬量》的机制不发动；本章预测其处置仍有多组等效解，且换一位医师换一张方仍可能等效。两篇在这一例上给出不同预测。**互补关系**：悬量管介入的**入口**，消参管介入的**归因**；一个体系可以入口通畅而归因永远做不出来。

《欠返》（《治未病立账学》第五章）（未病如何被识别与正当干预，同三学科）。那一篇处理的是返程未完成的那一段，本章处理的是内部量的不可分。**判决性对照：**一个返程完整（欠返率为零）的康复过程，那一篇的机制不发动；本章预测该过程中各项干预的独立贡献仍不可辨识。

四篇合起来：一篇记漏了判定线的移动，一篇记漏了未完成的返程，一篇记漏了无归属的风险量，本章记漏了被吸收的参数。四种漏记可以同时发生，且叠加时后果相乘而不是相加。

十五·本章不适用的地方

第一，可辨识的对象。血糖、血压、病原体载量、影像上的病灶尺寸。这些落在第①格，还原对它们完全有效，本章的机制不适用。若有人用本章去为拒绝检测辩护，那是误用。

第二，可辨识而不可控的对象。落在第②格的那些——年龄、基因型、既往暴露。本章关于「控制不需要辨识」的论述对它们没有意义，因为那里根本没有控制通道。

第三，急症与大扰动情形。如第十二之三节所述，不可辨识的控制扰动与工况迁移时会显著失效。本章不主张在急症中以经验控制律替代辨识；恰恰相反，本章预测那正是它最容易出事的地方。

第四，消参谱上靠近可辨识一端的那些成分。如第十二之二节所述，消参是程度问题。一套判定规则里可辨识的那部分，仍然应当被当作可还原的对象去研究，而且那正是还原研究应当集中火力的地方。本章不主张停止寻找物质基础，本章主张在开始寻找之前先算一次消参率，把找得到的与找不到的分开报告。

第五，本章管不了的一个洞：一个经验控制律是如何在没有辨识的情况下被积累起来的？本章说清了它一旦存在会怎样运转、为什么交不出参数，但没有说清它的形成机制——数亿次闭环调整如何在没有参数的情况下沉淀为可传授的规则。本章手头的材料不足以回答，姑且交出去。

十六·怎样才算本章错了

十六之一·最强的那个竞争解释

不挑好打的靶子。本章面对的最强竞争解释是这一句：

其实不就是样本量不够、人群太杂吗？等队列够大、分层够细，自然就收敛了。

它能解释本章引用的几乎全部现象：跨研究指标不重叠（样本不足、人群不同）、区间窄而彼此不重合（各自的人群不同）、维度增加而一致性不升（分层还不够细）。它不需要任何新概念，而且它是七十年里的主流处方。

要证明本章不是它的一个花哨说法，必须拎出一个**只有本章的机制成立才会出现、而竞争解释预测不出的变量**，并在控制掉竞争解释之后仍要求它成立。

那个变量是**消参率**。它有一个决定性的性质：**它与样本量、与人群构成完全无关**，它只由模型结构与观测方案的共现结构决定，在采集任何数据之前就能算出。竞争解释对它不作任何预测；本章预测它决定跨研究离散度的下限。

两个预测在同一份材料上给出可分辨的读数，因此可裁决。

十六之二·正式的证伪条件

条件一（主条件，机制证伪）：取同一制度环境内已发表的至少六个证候判定模型或群体药代模型，对每一个算出消参率（按第九之一节三条规则），并计算其跨研究参数估计的离散度。**控制样本量、技术世代与人群异质性指标之后，若消参率与离散度之间不存在剂量—反应关系，则本章的核心机制为伪——届时应把不收敛归因于样本与异质性，本章多此一举。**

条件二（激励，本章最反直觉的一条）：在同一体系内取规范化程度不同的至少六家机构，统计其入判项的两两完全共现对数。**若规范化程度与共现对数之间不存在正向关系，则本章关于「规范化即激励削减」的论断为伪。**这一条的数据在各机构的病历系统里已经存在，不需要新的采集。

条件三（轮次形态）：本章预测的不是一次先后，是一个循环：**观测结构先动（新增独立可变的观测通道）→可辨识组合清单在若干年内跟上→新的可辨识清单回写下一版判定规则里哪些项算独立入判项→下一轮由判定规则一侧先动，逼着观测结构改。**若在六次以上的标准修订与观测手段更替中，发起处始终在同一侧，则本章关于发起权轮转的部分为伪，本章退化为一**条单向因果说法**。

条件四（能稳住的标志物）：本章预测，历史上真正稳住的那些证候相关指标，其出现在时间上紧跟**观测通道的扩容**（新的、能独立变化的测量手段），而不紧跟统计方法的改进或样本量的提升。若编年后发现稳住的指标与统计方法改进同步而与观测扩容无关，则本章错。

条件五（扰动脆弱性）：本章预测消参率高的体系在工况迁移下的性能跌落更大。若在多个可比的临床或工程体系中，消参率与「条件迁移时的效果衰减幅度」无关，则第十二之第三节所依赖的那条削减不成立，本章关于急症弱势的解释作废。

条件六（共同因素判决）：若在心理治疗研究中严格控制共同因素之后，流派间效果差异显著出现，则本章关于「多组等价解」的读法错，共同因素论对。

条件七（可解释性）：若不同归因方法对同一模型的不一致程度与该模型的消参率无关，则本章对可解释性现象的解释错。

条件八（编码可复现）：若两组独立分析者按第九之一节三条规则对同一版判定标准与同一批病历计算消参率，所得结果差异超过该领域可接受的范围，则消参率不是一个可用的读数，本章的操作部分作废。

十六之三·样本纪律

条件一与条件二涉及跨单元比较，因此必须写死：

- **样本量不少于六个单元，且必须在同一制度环境内取。**可用的单元：同一部委下辖的多个病种判定标准、同一学会的多个版本、同一省份的多家三甲医院、同一药物的多项群体药代研究。
- **严禁用两个国家或两个地区充数。**「甲国如此、乙国如彼」在文化、监管、支付、记录习惯上处处不同，所谓控制变量在方法上站不住。它可以作启发式例证，**不能充当证伪检验。**
- 若确实只能跨国比较，必须在正文中写明这是启发式例证，并另给一个同质多单元的设计顶上。

十六之四·本章可以被分层引用

- **第一层（最强、最易倒）：**消参作为一类有本体地位的量，即第④格与第③格有实质差别。它被条件五（扰动脆弱性）直接检验——若被吸收的那份在任何情形下都不单独起作用，它就只是冗余，第④格应当并入第③格。
- **第二层（分析工具）：**可辨识—可控两根轴与四格。即便第一层被推翻，这张表仍可用于给任何一个研究纲领定位，指出它把资源投在了哪一格。
- **第三层（最稳、最便宜）：**一条不依赖前两层的经验主张——**中医证候判定标准的消参率从未被计算过一次，而它可以在不采集任何新数据的情况下算出**

来。这一条不需要接受本章的任何理论，它只是指出一件没有人做过、成本极低、而结果会直接影响资源配置的事。

说明本章可以被分层引用，是诚实，也是保护。

十六之五·若本章被证伪，会学到什么

若条件一失败——消参率与跨研究离散度无关——那么七十年的不收敛就确实是数量问题，而这将是一个有价值的结论：它意味着继续扩大队列是对的，只是还不够大。这会把资源留在现有方向上，但至少是**有理由地**留在那里，而不是靠惯性。

若条件二失败——规范化与激励贫乏无关——那么规范化就是无代价的好事，应当继续推进。这同样是一个值得知道的结果，因为本章最尖锐的实践含义正建立在它上面。

十七·这个判断被采用之后会怎样反噬

反噬一：把消参率写进评审，最省事的达标办法是删参数。一个研究者若被要求把消参率降到某个水平，最快的办法不是扩容观测——那要钱、要时间、要伦理审批——而是把不可辨识的参数从模型里删掉，或者施加一条便利性约束把它固定住。删完之后消参率是零，模型干净，评审通过；而被删掉的正是第④格里那些在控制上起作用的量。**降低消参率的最快路径，是把有用的东西删掉。**

反噬二：报告共现结构，最省事的达标办法是把模板写松。由于消参率的大小取决于规则三对有效观测通道的认定，任何机构都可以通过在记录模板里增加大量很少填写的可选项，把纸面上的通道数做大，从而把消参率做小。**规则一旦成为考核依据，规则本身就成了被优化的对象。**

反噬三：把「不必辨识也能控制」讲出来，最省事的接收方式是拒绝一切还原研究。「反正说不清，管用就行」这句话可以被用来拒绝任何机制研究、任何安全性评价、任何不良反应归因。第十二之一与十二之二节的两处削减正是为了挡住这个方向，但削减写在论文里，口号传得比论文快。

出口在哪里？我没有找到。任何提高可辨识性的做法，要么扩容观测（贵，且与规范化的方向相反），要么删参数（毁掉控制），要么接受组合层的理论（放弃拆解，而拆解是当前全部资源的去向）。三条路各有各的坏，本章能做的只有把它们写在这里，让采用者知道自己在选哪一种。

伦理边界另有三条，与第九之四节并列：

其一，本章的任何部分都不得被用作规避安全性评价的理由。**不可辨识与不必负责是两件事**：一个不能被分解归因的处置，其整体的安全性仍然必须被整体地评价，而且正因为不能分解归因，整体评价的标准应当更严而不是更松。

其二，本章不得被用作拒绝对照试验的理由。多组等价解的存在恰恰要求在**整方、整体层面**做严格的对照，而不是取消对照。

其三，在急症与快速恶化情形中援引本章的，应当同时援引第十二之三节——那一节明确预测了这类情形正是不可辨识控制最容易失效的地方。**只引前半段不引后半段的引用，是误用。**

十八·本章自己落在哪一格

一条关于「有些量说不清而管用」的判断，若不能用在自己身上，它就是自封的。

本章提出的消参率，按本章自己的两根轴判定，落在第①格还是第④格？表面上是第①格：它是纯结构计算，规则公开，可复算。但仔细看第九之一节的规则三——有效观测通道要从真实记录的共现结构中统计——这一步引入了大量分析者裁量：多少次算「始终」、多低的独立变化率算「不独立」、模板项与记录项如何对应。**这些裁量本身构成一组不可辨识的参数：不同的分析者会给出不同的消参率，而每一个都能自圆其说。**

也就是说，**本章的读数自己就是一个消参率不低的对象。**这不是修辞，它有具体后果，本章必须认下来。

后果一：本章不能靠自己算一次来证明自己。若只有本章作者算出高消参率，那与七十年里每一项得出自己那个估计值的研究没有区别。本章的第八条证伪条件（两组独立分析者的一致性）因此不是附带条款，是本章成立与否的第一道门。

后果二：本章给自己设了一条时限。若到二〇三六年，没有任何一组独立分析者能按本章公布的三条规则在同一版判定标准上算出彼此一致的消参率，那么按本章自己的判据，消参率就是第③格的东西——既不可辨识也不可控——应当被丢弃，本章的第一层与第三层一并撤回，只保留第二层的分类工具。

后果三：本章所处的位置正是它所描述的那一格。学术评审也是一套控制律：它每天在做的事是把稿件驱动到「可发表」这个目标状态，而它对「什么使一篇文章有价值」的内部参数，从来没有被唯一确定过——不同评审者给出不同的、各自自治的解。本章若被接受，多半不是因为它对，而是因为它落在了某一组等价解的可行域里；本章若被拒绝，也未必是因为它错。这句话对本章不利，但按本章自己的判断，它必定成立。

十九·结论，与一条写死日期的赌注

七十年的还原研究没有失败在技术上，也没有失败在样本量上。它一直在用一条先验地过不去的合格线，去评价一个从未走过那条通道的体系。可辨识与可控是两个互相独立的结构性质：一套判定规则可以在给定的观测结构下永远分不出内部的量，同时可靠地把系统驱动到目标状态。不收敛不是数量问题，是结构问题；有效不是运气，是因为控制不需要辨识。

承认这一点的代价是承认一类量的存在：它在辨识账本上被并入组合、因而不是一样东西，在控制账本上每天单独起作用。承认它，要花的钱是扩容观测通道；不承认它，要花的钱是每一代人重做一遍七十年，而这笔账已经付过七次。

最省事的第一步不需要任何理论上的让步：算一次。取现行的一版证候判定标准，取一批脱敏病历的入判项共现结构，做一次结构可辨识性分析，把消参率报出来。不需要新的病人，不需要新的伦理审批，不需要同意本章的任何一条判断。七十年里，这件事没有被做过一次。

给出一条写死日期的赌注：

到二〇三四年十二月三十一日，至少有一份中医证候相关的临床研究报告规范、行业指南或标准修订说明，要求在报告任何参数估计或权重之前，先报告该判定模型在所用观测方案下的结构可辨识性分析结果。

什么不算命中，一并写死：

1. 该要求只出现在方法学论文或综述中，未进入任何规范、指南或标准文本，不算命中。
2. 只要求报告数据层面的共线性诊断（方差膨胀因子、条件数一类），不算命中——那是数据层的性质，会随样本改变；本章要求的是结构层的分析，与数据无关。

3. 只要求「讨论模型的局限性」而不要求给出可辨识性的判定结果，不算命中。
4. 只在药代动力学或药理模型中要求，而不涉及证候判定模型的，不算命中。
5. 该条款由本章作者或其直接合作方参与推动写入的，不算命中——赌的是这件事会自己发生。

若到期未兑现，本章关于「这个分析的缺席是结构性的」这一判断得到反面证据：那说明它不是一个被制度性遮蔽的位置，只是一件还没轮到去做的事，本章对其结构性的估计过高了。

注释

〔一〕本章所称「观测结构」，指从原始观察到判定输入之间的全部通道及其在实际记录中的共现约束，包括记录模板、必填项设置、采样时点、以及执行者的记录习惯。它不包括判定规则本身。

〔二〕本章所称「控制通道」，指从判定结果到可执行处置之间的映射，以及该映射在反复使用中修正的机制。它不要求任何关于内部机制的知识。

〔三〕第九节的消参率与统计学中的「模型不可识别」在数学上同源，但报告习惯不同：统计学通常在遇到不可识别时施加约束并在方法部分一句带过，本章要求把施加约束之前的原始状态报告出来。**两者的差别不在数学，在于报不报。**

〔四〕第十三节对各既有说法的表述力求公允，未作简化；若有失当，责任在本章。

〔五〕第十六节条件一所需的编年与建模材料，目前尚无公开的整理版本，这是本章没有自己完成该项检验的原因，也是本章把它列为第一条证伪条件而非已完成证据的原因。

第九章 同据

——判定所依据的与验证所依据的是同一份信息：一类被自己的结果反向选出、因而没有回路外支点的分类

提要 「证」自二十世纪五十年代进入实验室以来，七十年间跨越六七代技术，始终没有收敛为一组稳定的、跨研究可复现的指标；与此同时，以辨证为前提的处置持续显示疗效。本章提出第三种解释，它不诉诸对象边界的漂移，也不诉诸内部参数的不可分：

「证」是一类被自己的结果反向选择出来的分类，其判定所依据的信息与其验证所依据的信息是同一份。要还原它，必须先找到一个回路之外的支点，把效据从判据里剥出来；而剥出来之后剩下的那一部分，不再是原来那个东西。它有效，是因为这条回路每天都在做筛选——任何在效据上不兑现的判定与处置的配对会被逐出体系，剩下的必然兑现。于是有效性的性质必须被重新说清：**它是回路内有效，是真实的、可重复的、对病人有价值的；但它不自动传递到回路之外的硬终点。**本章取细胞学的克隆选择、工程学的失效反选规范、管理学的例程选择保留三条互不相容的单因驱动主张相互顶撞，用免疫学自己在胸腺阳性选择与自身组织相容性限制里已经掌握、却从未往这个方向使用的一件事实——**选择的判据有一部分由被选择者所在的系统自己提供**——推翻三家共同假定的「选择的判据在被选择者之外」。由此得到一类在三家账本上只有负号而没有名字的东西：**判据与效据重合的那一份信息**，本章称之为**同据**。给出跨领域量纲相同的读数**同据率**：判定依据项集合与效果判定依据项集合的交集，占判定依据项集合的比例；并给出一句不含情态词的判据：这一次判定所用的观测项，与判定这一次有没有效所用的观测项，有几项是同一项？全文最锋利的一条可裁决主张是：**控制住指标的主客观属性之后，同据率仍与报告效应量之间存在剂量—反应关系；而同据率高的体系，其硬终点效应接近零但并不恶化——这一点使它与指标失真的通行诊断分得开。**

一·引言：还原需要一个支点，而这个支点从来没有被找过

七十年的证候客观化研究有一个从未被明说的作业形式：**先由中医专家按现行标准把病人判为某证，再在这组人与对照组之间找差异指标。**这个形式假定了一件事——判定这一侧是可以被当作参照来用的。找到的指标之所以被称为「该证的物质基础」，是因为它与判定对上了。

那么判定本身凭什么算数？

在纸面上，答案是「凭标准」。标准写着哪些表现、什么权重、如何组合。但标准里的条目又是从哪里来的？往上追，追到的是历代医家的经验；再追一

层，追到的是**这些条目所对应的处置在临床上兑现过**。一个证名之所以留在体系里，从来不是因为有人先测出了它，而是因为按它施治的方子管用。**判定的正当性最终落在疗效上**。

这句话本身不新，它是中医方法学里公开讨论了几十年的「以方测证」。新的是接下来这一步：**那么疗效又是用什么判定的？**

翻开任何一份中医临床研究的方案，答案清清楚楚：证候积分。判定这个人是不是「脾肾阳虚」，用的是畏寒、便溏、腰膝酸软、舌淡、脉沉细；判定治疗有没有效，用的是这几项各自的分值下降了多少、总分降了百分之几。**判据与效据，是同一批条目**。

于是形成了一个闭合的回路：条目定义了证，证决定了处置，处置的成败又用同一批条目来判，判出来的结果再回过头确认这些条目选得对。

这个回路在中医内部通常被辩护为「符合中医自身规律」，在外部通常被批评为「循环论证、疗效不可信」。两种说法都太快了。本章要说的是：

这个回路既不是需要辩护的特色，也不是需要修补的缺陷，它是一个筛选装置。而一个筛选装置的产物有两个必然的性质：它必然在筛选判据上兑现（否则早被筛掉了），它也必然无法用回路之外的量来还原（因为它的定义里含着回路本身）。

七十年的还原研究之所以做不成，不是因为对象太复杂、样本太小、人群太杂、技术太粗。是因为**还原需要一个支点——一个独立于被还原对象的参照——而在这个回路里，这个支点不存在**。每一次「找到了标志物」，其实是找到了一个与判定条目共变的量；而判定条目又是被疗效条目校准的，疗效条目又就是判定条目。整条链上没有一处接地。

这个说法要立住，需要三件事：需要说清是什么在驱动这种回路的闭合；需要给出一个别人能在别人的材料上清点出来的读数；需要给出一个能让本章栽跟头的观察。这三件后面逐一交出。

本章还必须处理一件更要紧的事，否则它就是一篇拆台的文章：**如果判据与效据同源，那病人的好转是假的吗？**不是。本章的核心区分是**回路内有效与回路外有效**。回路内有效是真实的——病人的畏寒确实减轻了，便溏确实好了，这些是他自己的生活质量，不是幻觉。它只是**不自动**意味着心血管事件下降、住院率下降、生存期延长。批评者说「那不是真有效」，辩护者说「病人确实好了」，

两边都对，因为他们说的是两种有效性，而现行的话语里没有把它们分开的词。本章提供这个词。

二·把「以效验证」这四个字拆开

「以效验证」在日常语言里混着三件很不同的事，不拆开就没法讨论。

第一件：疗效作为发现的线索。一个方子管用，提示其中可能有活性成分，于是去找。青蒿素是这条路的典范。这一件完全正当，而且是回路外的——检验活性成分的判据是原虫清除率，与「病人自述好些了」是两回事。

第二件：疗效作为分类的筛子。一个证一方配对若长期不兑现，会被弃用。这一件也正当，它是任何实践体系都必须有的淘汰机制。要紧的是问：**筛子的孔是用什么材料做的？**若用的是回路外的硬终点，那么筛出来的东西可以被回路外的语言描述；若用的是与判定同一批条目，那么筛出来的东西只能被回路内的语言描述。

第三件：疗效作为分类的定义组成部分。当「有效」这件事本身被定义为「那几个判定条目的分值下降」时，判定与验证已经不是两个环节，而是同一个操作的两次执行。这一件不是不正当，但它使**还原在原理上无处下手**。

三件事在文献里常常被混在一起，因为它们的表述形式一样，都是「用疗效来说明证的合理性」。区分它们的方法很简单，就是问一句：**判据的条目与效据的条目，重合了几项？**

- 重合为零：第一件与第二件的良性形态。判定与验证互相独立，还原有支点。
- 重合为一部分：混合形态，最常见。
- 重合为全部：第三件。此时判定与验证是同一次测量，**还原没有支点**。

这个区分之所以从来没有被系统地做过，不是因为难做——它只需要把两份条目表并排放，数一数——而是因为**没有任何一份研究报告规范要求把这两份表分别列出来**。方案书里，诊断标准写在纳入排除那一节，疗效评价写在结局指标那一节，两节之间隔着十几页，从来没有人被要求把它们对齐。

三十年里，成千上万项中医临床研究执行了这个形式，而这个数一次也没有被算过。本章的第三层主张就是这一条，它不依赖本章的任何理论：**这个数可以在不采集任何新数据的情况下算出来，而它从来没有被算过**。

三·三种现成解释，以及它们共同站着的那块地

在引入三门外学问之前，先摆出本领域内最有力的三种解释。

三之一·复杂性说：整体状态无法被单一指标刻画

第一种解释说：证是多系统耦合的整体状态，任何单一指标只能抓到一角，需要系统生物学式的多维刻画。它的合理性在于对象确实复杂。

它的困难是：复杂性预测的是**慢收敛**，不是不收敛。而且它对本章提出的那个问题完全没有回应——无论刻画得多细，用来校准这套刻画的参照仍然是专家判定，而专家判定的正当性仍然落在疗效上，而疗效仍然用判定条目来测。**维度增加不会创造出一个回路外的支点。**

三之二·主观性说：判定者不一致

第二种解释把问题归给识别端的信度，处方是条目化、量表化、辅助判定系统。这条路的成绩是实打实的，一致性系数确实上去了。

它的困难是：**提高一致性并不触及回路**。让十位医师对同一个病人给出同一个证名，只说明他们用的是同一套条目；它不说明这套条目对上了病人身上任何一个独立存在的东西。**一致性衡量的是判定者之间的距离，不是判定与对象之间的距离**。而在一个判据与效据同源的体系里，把条目统一得越好，回路闭得越紧。

三之三·客观化未完成说：还没找到那个指标

第三种解释最朴素：只是还没找到，继续找。

它的困难是七十年的形势本身。更要紧的是：**这条路的每一次「找到」，其确认方式都是与专家判定的一致性**。一个指标之所以被认为找到了脾肾阳虚的物质基础，是因为它在被判为脾肾阳虚的人群里显著不同。这就把新指标接进了同一个回路，而不是把回路接地。**用回路内的参照去确认一个新变量，得到的仍是回路内的变量。**

三之四·三家共同站着的那块地

三家分歧不小：一家归给对象复杂，一家归给判定者不齐，一家说时间不够。但三家共有一条从不被说出的假定：

证的判定，是一件独立于它的处置结果的事。

复杂性说假定证在那儿等着被刻画，处置是后续的事；主观性说假定证在那儿等着被一致地识别；客观化说假定证在那儿等着被找到。三家争的是**怎么把它看清楚**，从没有一家问过看它的那把尺子是不是它自己的疗效磨出来的。

把这条假定念给三家听，三家都会说：这还用说吗？病人身上有什么，和后来我们治得好不好，当然是两件事。

这句「这还用说吗」，正是本章要动的地方。要动它，需要三条互不相容的、各自主张「反选就是全部」的强硬立场，把这块地照亮；以及一件出自这三家之一自己的材料，把它顶翻。

四·从三门学问各取一条驱动

选这三门，不是因为它们与中医像。选它们的唯一理由是：三家都主张「有效的适配不是被设计出来的，是被反选出来的」，而三家对**被反选的是什么**给出了互不相容的答案，且每一家都主张只有自己那条是决定性的。

四之一·细胞学：认得出什么，由选择走出来的路径决定

问题的形状：一个人的免疫系统能识别的抗原种类，远远超过基因组能编码的蛋白种类。这种识别能力从哪里来？

细胞学在这件事上的强硬主张是：**没有任何一种抗体的特异性可以从产生它的机制推导出来；特异性是选择的产物，不是设计的产物**。淋巴细胞在成熟过程中通过基因片段重排随机生成庞大的受体库，抗原进来之后，结合得上的那些克隆被扩增；扩增过程中受体基因发生高频突变，结合得更好的又被进一步扩增。这一整套过程被称为克隆选择与亲和力成熟。

这条主张的强硬之处在于它取消了「先有设计图、再有产品」的图像：**认得出什么，是被抗原环境走出来的一条路径，不是被写在基因组里的一份清单**。

它的动力式是：

淋巴细胞当下的结合表现，与它所处的抗原环境之间的矛盾，逼出增殖与突变走哪一条路径；新的路径回过头改写了这个细胞能显露成什么、以及它此后会遇到什么样的环境。

时序读数：先动的通常是抗原环境；克隆扩增与超突变在数天到数周内跟上；记忆库的重塑滞后数月，而库的整体结构要经年才改变。跟不上的永远是库的整体结构，因为它承载着一生的历史。

这一家自曝的一处：先天免疫的模式识别受体是基因组直接编码的，特异性固定，不经任何选择，而它对绝大多数病原的第一道防线恰恰靠它。也就是说存在一部分识别能力**不是反选出来的**。细胞学知道这件事，把它写在先天与适应的二分里，从不把它读作「反选不是唯一来源」的证据。

四之二·工程学：什么算安全，由失效史反过来写成

问题的形状：一根梁该做多厚？教科书上有公式，公式里有一个安全系数。这个系数是多少？从哪里来？

工程学的强硬主张是：**工程规范里的数字没有一个能从第一性原理推出来；它们全部是失效历史的沉淀。** 加压客舱窗角的疲劳裂纹改写了整个疲劳设计体系；一座桥在风中扭转崩塌改写了风振设计；军机与客机的几次断裂改写了从安全寿命到损伤容限的整个范式。工程界有一句流传很广的说法：规范是用血写成的。这不是修辞，它描述的是规范条文的实际来源——**每一条许用值、每一个检验间隔、每一项强制检查项目，背后都有一次具体的失效。**

它的动力式是：

结构在服役中实际显露出的表现，与设计所走的那条路径之间的矛盾，逼出设计所依据的整个条件场改变——许用应力、载荷谱、检验制度、认证基础；新的条件场回过头改写了什么算一次合格的表现、以及此后该怎么设计。

时序读数：先动的通常是实际表现，一次失效；规范修订在两到五年内跟上；教材与工程教育滞后十年以上；而已在役的存量装备的改造滞后最久，常常永远做不完。

这一家自曝的一处：计算力学确实成功地预测过从未发生的失效模式——复合材料的分层扩展、某些屈曲模式，是先被算出来、再被实验证实的。也就是说存在一条**不依赖失效史**的路径。工程学把它写在计算力学的成就栏里，从不把它读作「反选不是唯一来源」的证据。

四之三·管理学：能做成什么，由被保留下来的例程显露出来

问题的形状：一家公司为什么能做成某件事，而另一家做不成？把它的流程文件全部复制过去，为什么复制不成功？

管理学有一支的强硬主张是：**组织的能力不是被设计出来的，是被选择保留下来的例程；没有任何人能说出一条例程的全部内容，而它照样被可靠地执行、被有限地复制。**演化经济学把这条说死了：变异、选择、保留三个环节构成组织能力的实际来源，而所谓战略选择在很大程度上是事后的合理化叙述。

它的动力式是：

组织实际走的那条路径，与它所处的环境之间的矛盾，逼出它显露成什么样的能力与绩效；新的显露回过头改写了它下一步会走什么路径、以及它被放在什么环境里。

时序读数：先动的通常是环境；例程的变异—选择—保留在数年内跟上；正式制度文本滞后最久，常常在实践早已改变之后才追认。

这一家自曝的一处：刻意学习与成文化确实提升了例程的复制成功率——把做法写下来、做后复盘，能让能力更可靠地传递。也就是说，例程并非只能靠盲目的选择保留。管理学把它写在刻意学习的文献里，从不把它读作「反选不是唯一来源」的证据。

四之四·三家不能同真在哪一点上

三家都说「反选」，而被反选出来的那一端互不相同：细胞学说被驱动的是**路径**（增殖与突变走哪条道），工程学说被驱动的是**条件场**（规范与许用值），管理学说被驱动的是**显露**（能力与绩效）。

这不是三个角度，是三条不能同真的主张。若细胞学对，则工程界每一次失效之后改的应当是**设计的做法**而不是规范的数字；若工程学对，则组织在环境压力下改的应当是**它据以判断合格的标准**而不是它的能力表现；若管理学对，则免疫系统在抗原压力下变的应当是**它当下的应答强度**而不是受体库的路径。三家在各自领域里都能举出支持自己的证据，而三条不能靠判谁对化解。

唯一的出路，是承认一件三家都没说的事。要说这件事，需要材料，而材料必须来自三家之一自己。

五·翻转发生在哪一处：免疫学自己的那件材料

材料在细胞学手里，而且是免疫学最基本、最有名的一段结果。

淋巴细胞在胸腺里成熟时要过两道关。第一道叫阳性选择：**只有那些对自身的组织相容性分子加自身肽有中等亲和力的细胞才被允许存活**，完全不结合的会因缺乏存活信号而死去。第二道叫阴性选择：结合得太强的会被清除，以免攻击自身。

这里有一件事，它在免疫学里是常识，被写进每一本教科书，而它从来没有被读作本章要读出的那个意思：

判据的一部分，是由被选择者将来要服务的那个系统自己提供的。

一个 T 细胞将来要识别的是外来抗原，但它得以出生的门槛，是它对**自身**的分子有多大亲和力。识别外来的能力，是被自身的背景筛出来的。这就是所谓的自身组织相容性限制：**T 细胞不能单独识别抗原，它只能识别与自身分子一起呈现的抗原**。识别的判据里已经写着自身。

更进一步：一个人的自身肽库塑造了他的淋巴细胞库，而这个细胞库反过来通过清除、调节与呈递，塑造了这个人此后会呈现哪些自身肽。**判据与被判者互相构成**。

在免疫学自己的问题结构里，这段材料回答的是另一个问题：为什么 T 细胞的识别需要组织相容性分子参与，以及自身耐受是怎么建立的。判据在不在被选者之外，与那个问题无关，所以没有人往那儿看。

这件材料一旦被摆到这个位置上，三家的共同假定就顶不住了，而且顶不住的原因不是有人反对，是反例出自三家中最讲究「**外部环境说了算**」的那一家自己：那家主张「抗原环境决定一切」的免疫学，自己的核心机制里存放着一套**判据由内部提供**的选择过程。

于是可以说出那句被共有前提挡住的话：

选择的判据不必在被选择者之外。当判据的一部分由被选择者所在的系统自己提供时，选择仍然照常进行、照常筛出真正有效的适配——但筛出来的这份有效，只在这条回路之内成立；而它的内容，永远不能用回路之外的量还原出来，因为它的定义里含着回路本身。

要立刻补一句，否则这个判断会被读成虚无主义：**回路内选出来的东西可以在回路外真的管用**。用自身背景筛出来的 T 细胞库，确实能对付真实的病毒。这

一点极其要紧，它是本章与「循环论证所以无效」这种廉价批评之间的全部区别，第十二之一节会把它变成一处对本章的削减。

三家看见的是这个循环里的不同段落：细胞学看见「环境先动、路径跟上」那一段，工程学看见「表现先动、规范跟上」那一段，管理学看见「环境先动、能力显露跟上」那一段。三家都没错，三家都只看见了自己那一轮。

六·承重判断：同据

现在可以把判断写出来。

「证」不是一个尚未被多维刻画的复杂状态，不是一个尚未被一致识别的主观判定，也不是一个尚未被找到的客观指标，而是一类被自己的结果反向选择出来的分类，其判定所依据的信息与其验证所依据的信息是同一份。

判据与效据重合的那一份信息，本章称为**同据**。

这个名字要担的意思有两层，缺一层就不成立。

第一层：它不是偏倚，它是一份东西。在现行的任何一套方法学里，判据与效据的重合都被记为**负项**：工业与组织心理学叫它准则污染，测量学叫它共同方法偏差，工程验收叫它自证信号，实验生物学叫它循环标记。四个名字，同一种处理——**减掉它，或者用独立来源替换它**。也就是说，在所有账本上，它只有一个负号，没有一个名字。它不是被记录得少，是**它压根不算一样东西**。删掉这个读数，它不是变得难测，是不存在。

第二层：它是筛选装置的工作面。减掉它之所以不成，是因为这类分类正是**靠它被筛出来的**。一个证一方配对之所以还留在体系里，是因为它在那几个条目上兑现过；把那几个条目从疗效判定里剔除，筛子就没有孔了，此后既筛不出新的配对，也无法解释旧的配对为什么在这里。**减掉同据，被减掉的不是噪声，是这套分类赖以存在的机制。**

有了这两层，开头那件怪事就有了单一的来源。

它不可还原，是因为还原需要一个回路外的支点，而这个支点不存在。还原的标准形式是：找一个独立的量 X，证明它与证有稳定关系。但「证」是用条目集合 C 判定的，而 C 的正当性来自它在疗效判定上兑现，疗效判定用的也是 C。于是 X 只能被拿去与 C 比对——**比对的结果无论多好，都只是 X 与 C 的关系，不是 X 与「证」的关系**，因为「证」除了 C 之外没有别的落脚处。七十年里每一次

「找到了标志物」，找到的都是与 C 共变的量；换一版 C，就换一批共变的量。
这不是精度问题，是链条没有接地。

它有效，是因为这条回路每天都在做筛选。数千年、数亿诊次的实践里，任何在 C 上不兑现的配对都会被逐出；留下来的必然兑现。这份兑现是真的：病人的畏寒确实减轻，睡眠确实变好，这些是他自己的生活，不是统计假象。**但这份兑现的判据是 C，因此它的传递范围也止于 C 所覆盖的那些东西。**

于是必须把有效性拆成两个：

- **回路内有效：**在筛选所用的那份判据上兑现。这一份是被保证的，因为不兑现的早已被筛掉。
- **回路外有效：**在筛选从未用过的判据上兑现——死亡率、住院率、复发率、并发症、生存期。这一份**没有任何机制保证它**，它可能有，也可能没有，取决于回路内判据与回路外目标之间碰巧有多大的共变。

七十年的争论之所以谁也说服不了谁，是因为双方一直在用同一个词说两件事。批评者拿硬终点说话，辩护者拿病人说话，两边都诚实，两边都对。**本章不裁决谁对，本章给这两件事各一个名字，并给出一个可以清点的读数来分开它们。**

为什么这三家自己到不了这一步。不是疏忽。细胞学的核心问题是免疫系统如何工作，它没有理由去问「选择的判据在不在被选者之外」，因为在它的问题里判据是什么无关紧要，只要选择能跑。工程学的核心问题是不出事，而它的规范体系必须被当作外部权威才能履行监管职能——一旦承认规范是被自家失效史反选出来的、且合格判定与失效判定共用同一套量，整个认证体制的正当性就要重新论证；停在「规范是经验的沉淀」这句谦逊话上，对它是最安全的。管理学的核心问题是能力从哪来，它已经承认了例程不可完全言明，再往前一步就要问「用什么判断一条例程是好的」，而那会把绩效评价整个掀翻。**一个领域看不见什么，通常不是它的缺陷，是它的问题结构的影子。**

七·两根轴与四格

一个名字不能分东西，两根轴才能。

- **轴一：判据是否独立于被判对象。**两端是「判据外生」与「判据与对象共源」。

·轴二：这套配对是否被持续筛选。两端是「无持续筛选」与「有持续筛选」。

	无持续筛选	有持续筛选
判据外生	② 未经检验的设计：算得出来但没跑过实战的规范；上市前的新药；纸面上的组织流程。判据是硬的，但还没有被结果筛过。	① 回路外有效：抗生素对细菌（判据是菌落清除，与用药决策无关）、外科手术对骨折复位、疫苗对发病率。判据外生，筛选持续，还原有支点。
判据共源	③ 自说自话：无人使用、无人淘汰的自评体系；只有一次的内部评估。既没接地也没被筛过，一眼可见地无价值。	④ 同据：「证」与证候积分；用自身背景选出的淋巴细胞库；用自家失效史写成又用自家规范判合格的工程体系；组织自己定义并自己测量的核心能力。改得动，说不清，而且说不清的原因不在测量，在参照。

四格必须落到具体的事，否则这张表只是修辞。

第①格是循证医学的理想格，也是还原真正成功的地方。它有一个可核查的签名：判定用的量与验收用的量不是同一批，因此可以被第三方独立复核。

第②格常被误当作第①格。一条从第一性原理算出来、逻辑无懈可击而从未经过服役检验的规范，与一条被几十次事故打磨过的规范，在文本上看不出差别。这一格的风险在工程史上反复出现。

第③格是坏的，而且一眼可见地坏。要紧的是：第③格与第④格在任何一次单独的评估里看起来一模一样——两者都是自己判自己。区分它们的唯一方法是去看有没有真实的淘汰：这套配对里，有没有东西曾经因为不兑现而被逐出？若从来没有，它就是第③格。

第④格是本章要打的那一格。它与第③格只差一个筛选，与第①格只差一个支点，而现行的评价体系只有一根轴——它把第④格与第③格一起判为「不可信」，把第④格与第①格一起写进「有效率百分之九十二」。两个方向都错，而且错得相反。

八·为什么第④格能通过全部形式审查

签名一：短期指标全面改善，而且改善是真的。

第④格里每一项常规指标都好看。有效率高——因为疗效用的就是判定条目，而处置正是按这些条目选的；一致性高——条目统一；病人满意度高——那些条目本来就是病人最在意的主观感受；论文顺畅——阳性结果易得。

要紧的是：**这些好看的数字没有一个是假的**。没有人造假，没有人违规。第④格的全部麻烦在于，所有这些真数字合起来仍然不告诉你回路外发生了什么。

签名二：失效不产生任何信号。

设想一次真正的失效：一套证—方配对在证候积分上稳定兑现，而在硬终点上是零效应，甚至因为延误了别的处置而略有损害。这次失效会在哪里被记录？

不会在临床记录里——每个病人都判了证、开了方、积分下降、记了显效。不会在研究报告里——主要结局达成。不会在指南里——指南采纳的正是这些研究。不会在标准修订说明里——修订记的是条目增删，不记「本次修订后判据与效据的重合项数变化」。

判据与效据的重合，不出现在任何一份报告的任何一个字段里。这不是有人隐瞒，是没有这个字段——因为方案书里诊断标准与结局指标分属两节，从来没有人被要求把它们对齐。

签名三：自我加固，越正规化越严重。

这一条最容易被检验。规范化做的是什​​么？统一术语表、条目化判定、量化疗效评价、制定统一的证候积分量表。这些动作全部把判据与效据往同一份术语表上收。**术语表统一的过程，同时是判据与效据重合度上升的过程**。

而重合度越高，报告出来的疗效数据越好看——因为处置正是按这些条目挑的；数据越好看，越容易立项、越容易进指南、越容易成为下一批研究的参照。**做得越好，回路闭得越紧，而所有的表都显示做得越来越好**。

这一条不是推测。任何一批已发表的中医临床研究，把它的诊断标准条目表与疗效评价条目表并排放，数一数重合项，就能算出来。**数据已经存在，不需要新的采集**。

九·一个可以清点的读数：同据率

九之一·定义与编码规则

同据率＝在一次判定与其效果评价中，两者所依据的观测项集合的交集大小，占**判定依据项集合**大小的比例。

它是一个比例，无量纲，跨领域可以并排比较。它有一个别的读数少有的优点：**它只需要两份文本**——诊断标准的条目表与疗效评价的条目表。不需要病人，不需要样本，不需要伦理审批。

清点它需要三条编码规则，缺一条这个数不可复算：

- **规则一：什么算一个观测项。**在标准文本中被单独列出、可被单独记录取值的项。同一项在两份表中措辞不同而所指相同的（「畏寒肢冷」与「畏寒」），计为同一项；这一判定必须由两名独立编码者作出并报告一致性。
- **规则二：什么算重合。**两项在同一次就诊中由同一次观察行为获得，即为重合。**由不同观察行为获得的同名项不算重合**（例如判定时的患者自述与评价时的活动记录仪读数，虽同名而不重合）。这一条把疑难案例判给对手一侧。
- **规则三：以判定依据项集合为分母。**之所以不用并集或效据集，是因为本章关心的是「判定这一侧有多少落脚点被效据占用了」。若效据包含判据之外的项，那些项进入第二个字段而不进分母。

必须同时报告的**第二字段：回路外项数**——效据中不属于判据的观测项数目，且其中属于硬终点（死亡、住院、复发、并发症、失能）的项数。同据率为一而回路外项数为零的研究，与同据率为一而另附三项硬终点的研究，性质完全不同。**只报一个数是不够的。**

九之二·四个领域里的同一个比率

免疫学实验：判定一个细胞属于某亚群所用的标记，与判定该亚群功能实现所用的读出，之间的重合项数。某些活化标记既被用作分群依据又被用作活化程度的读出，此时同据率不为零。这在实验设计里被要求避免，做法是**换一个独立读出**——免疫学早就知道这件事，并把它写成了实验规范。

工程验收：判定设备处于某工况所依据的传感器通道，与判定该工况下控制性能是否达标所依据的通道，之间的重合数。增益调度系统里这个数常常不低，因为区判定与性能评估往往共用同一组信号；工程实践的处理是**加装独立的验收传感器**。

组织评价：组织诊断量表的条目与干预效果量表的条目之间的重合数。管理学与心理测量对此有成熟的警觉，并发展了多来源、多时点、多方法的设计来压低它。

中医临床研究：诊断标准条目表与证候疗效评价条目表之间的重合数。在大量已发表方案中，这个数接近分母的全部。据本章所知，**它从未被系统统计过一次。**

四处的量纲完全不同——标记、传感器通道、量表条目、四诊项——而比率的定义是同一个。要紧的是另外三家**都已经知道这件事、都已经发展了压低它的做法**，只有一家没有；而这一家恰恰是唯一把它当作体系特色来辩护的。

九之三·与既有主要指标正交的那个例子

例子是这样的：一项设计规范、随机、双盲、多中心的临床试验，报告某方对某证的证候疗效有效率为百分之九十二，主要结局达成，纳入指南。同一份方案里，诊断标准列了七个条目，疗效评价量表列了同样这七个条目——**同据率为一，回路外项数为零。**

有效率最高的那一批研究，同据率也最高，因为处置正是按这些条目挑出来的，而疗效又用同一批条目来测。两个数在方向上是同向上升的，而它们衡量的是相反的东西：一个衡量兑现程度，一个衡量兑现的范围有多窄。**任何现行的质量评价框架——随机、盲法、样本量、失访率——都不会碰到它。**一项试验可以在所有偏倚风险条目上都评为低风险，而同据率为一。

这就是本章这个读数与既有主要指标正交的证据：**质量评级最好看的时候，这个数最难看。**

九之四·这个数的伦理边界

同据率会被拿去评价研究、进而评价研究者与整个学科，因此三条必须写死：

其一，这个读数指的是设计，不是人，更不是疗效的真伪。高同据率不意味着病人没有好转，也不意味着研究者不诚实。它只意味着**这项研究没有提供关于回路外的信息。**把它读成「造假」或「无效」，是对本章的严重误用。

其二，不得用它作为削减既有服务的依据。回路内有效对病人是真实的价值——症状、睡眠、食欲、体力，这些就是生活质量本身。任何以「同据率高」为

由取消对症状改善的支付或供给，都违反本章的主张。本章要求的只有一件事：**把回路外的那一栏空着的事实公开写出来，而不是让它被有效率百分之九十二遮住。**

其三，据此作出不利评价的，必须公开三条编码规则并给出复核通道。同据率的大小取决于规则一对「同一项」的认定，任何据此作出的评价都必须允许被评价者按公开规则复算。

十·一句不含情态词的判据

这一次判定所用的观测项，与判定这一次有没有效所用的观测项，有几项是同一项？

这句话里没有「应当」「充分」「真正」「合理」。它问的是两份表的交集，答案是一个整数，由文本直接给出。

把它拿到五个场景里跑一遍，五个答案互不相同：

1. **一份中医临床试验方案**：诊断标准条目与证候疗效评价条目的交集数。答案常常等于诊断条目的全部，且这一点在方案书里从不出现，因为两份表隔着十几页。
2. **一份免疫学论文的方法部分**：分群标记与功能读出的交集数。答案通常是零或一——因为审稿人会问，这已经是这门学问的规矩。
3. **一份工业控制系统的验收报告**：工况判定通道与性能评估通道的交集数。答案居中，且往往因为加装独立传感器成本高而被容忍，报告里会写明。
4. **一份组织发展项目的评估报告**：诊断工具条目与效果工具条目的交集数。答案在使用同一套框架前后测的项目里等于全部，而在使用独立业务指标的项目里接近零。
5. **一份教育干预的评估**：判定学生需要该干预所依据的表现，与判定干预有效所依据的表现，交集数。答案在使用同一套量表前后测时等于全部。

五个答案里，至少有两个不能从本章的判断直接推出来：免疫学把这条写进了审稿规矩这件事、以及工程验收里独立传感器的实际配置比例——这两条只能查，不能想。

这五个场景合起来还提示了一件比命题本身更有用的事：在五个领域里，有三个已经把这个数按住了，做法各不相同（换独立读出、加装独立传感器、多来

源多时点设计），而没有一个领域是靠「取消筛选回路」做到的。**压低同据率的成熟做法全部是「加一个回路外的观测」，不是「拆掉回路」。**这一点直接决定了本章能给出的处方形状。

十一·在别的领域里能预测出什么

以下十二处，每一处都给出一个可以去查的预测。

一、中医临床研究文献：在同为主观结局的研究之间，同据率与报告效应量之间存在正向的剂量—反应关系；而在同据率相近的研究之间，效应量与盲法质量的关系较弱。这可用现有文献做一次元回归。

二、硬终点研究：以证候分层为依据的处置，在改用纯硬终点评价时，效应量的下降幅度应当与该研究的同据率正相关；而不以证候分层为依据的整方处置，其效应量下降幅度与同据率无关。**这是本章与「主观指标虚高」这个通行解释之间的分离线。**

三、免疫学：在自身肽库被人为改变的模型中（如改变呈递分子的等位型），成熟淋巴细胞库的特异性谱应当随之系统性改变，而其对真实病原的整体防御力**不必**同步下降。这一条支持本章关于「回路内选择仍可产出回路外能力」的那处削减。

四、工程规范：一条规范条文所依据的失效事件数，与该条文所约束的失效模式在此后的复发率之间，应当存在负相关；而**从计算得出、无失效史支撑的条文**，其所约束模式的复发率应当更高。这是可以在事故数据库里查的。

五、认证体制：合格判定与失效判定共用同一组量的认证体系，其认证通过率应当更高，而被认证对象在**认证未覆盖的失效模式**上的事故率不因认证而下降。

六、组织绩效：使用同一套框架做前后测的组织发展项目，其报告效果量应当显著高于使用独立业务指标的项目；而两类项目在**三年后的组织存活与财务表现**上应无差别。

七、绩效考核：一个考核指标若同时被用作诊断依据与验收依据，其达标率上升的速度应当快于被单独用作验收依据的指标，而与实质产出的关系更弱。

八、教育评价：使用同一量表前后测的干预研究，其效应量应当高于使用独立迁移测验的研究；而两类干预在**延迟迁移测验**上的差异应当远小于在即时量表测验上的差异。这在教育心理学里有大量现成数据可查。

九、机器学习的评测：在训练目标与评测指标高度重合的基准上，模型排名的提升速度快，而其在分布外任务上的表现与基准分数的相关较弱。本章预测这种脱节的程度与「训练目标与评测指标的重合度」正相关，且可以直接计算。

十、生态修复：以物种清单为修复目标又以物种清单为验收依据的项目，其验收通过率高，而生态系统功能指标（生产力、水文调节）的改善与验收通过率无关。

十一、临床指南的自引结构：一条推荐所依据的研究，若其结局指标取自该推荐所依据的诊断标准，则该推荐在不同人群中的效果方向应更容易不一致。

十二、中医方剂筛选：以证候积分为筛选判据的新药开发管线，其候选方进入以硬终点为主要结局的确证试验后的成功率，应显著低于以硬终点为筛选判据的管线。这是可以在注册数据库里追踪的，且直接关系资源配置。

十二·三处反过来给本判断加的约束

十二之一·免疫学削掉了「回路内有效不传递到回路外」

胸腺阳性选择用的是自身背景，而选出来的淋巴细胞库确实能对付从未见过的真实病原。这不是侥幸：**用自身背景作筛选判据，恰恰是特异性得以精细化的必要条件**——脱离自身背景的识别是粗糙的，而正是自身分子提供的稳定框架，使得对外来肽的分辨精细到单个氨基酸。

如果没有这一处，本章原本会写下的更强的一句是：「回路内有效不传递到回路外，因此证候疗效数据不能作为任何外部结论的依据。」这句话现在不能写了。

被迫承认的是：**传递是可能的，且传递的程度取决于回路内判据与回路外目标之间的共变程度**，而这个共变程度是可以被估计的，不是先验为零。于是本章的适用范围被削掉一大块：本章不能主张回路内数据无价值，只能主张**回路内数据不自动承载回路外结论，而这两件事目前在报告里是不分的**。这比原来的主张弱，也更可能是对的。

十二之二·工程学削掉了「反选出来的东西不可还原」

彗星号事故之后的几十年里，那些从失效史里长出来的经验系数**被还原了**——断裂力学给出了应力强度因子，疲劳裂纹扩展有了定量的扩展律，损伤容限从一套经验做法变成一门可计算的学科。也就是说，被失效反选出来的适配，在事后是可以被还原的。

如果没有这一处，本章原本会写下的更强的一句是：「被反选出来的分类原则上不可还原，继续寻找证的物质基础是徒劳的。」不能写了，而且被削掉的是本章的价值排序。

正确的表述是：**反选出来的适配可以被还原，但还原它所需要的，是一门尚未存在的中间层理论，而不是更多的分子层指标。**断裂力学不是把彗星号的经验系数还原成了原子间作用力，它是在**构件与裂纹这一中间尺度上另立了一套量。**对应到本题：证的还原若有出路，出路在**症状之间的关系结构与处置反应之间的定量规律这一中间层**，不在继续往下追分子。这一处**削减不只削弱了本章，还把本章的处方从「别找了」改成了「换一层找」。**

十二之三·管理学削掉了「条目化会毁掉它」

刻意学习与成文化的研究表明，把例程写下来、做后复盘，确实提升了能力的复制成功率。成文化不必然摧毁被反选出来的东西。

如果没有这一处，本章原本会写下的更强的一句是：「**条目化与量表化是回路闭合的元凶，应当退回到师承。**」不能写了。

被迫承认的是：条目化本身是中性的，**出问题的不是条目化，是把同一批条目同时放在判定端与验收端。**这个区分很要紧，因为它把本章的处方从一个不可行的方向（退回师承）改到了一个可行的方向（**继续条目化，但把两端的条目表分开建、分开报**）。这一处**削减因此同时是本章最有操作性的一处。**

十三·与最近的几种说法的分界

本章的判断听上去像好几种已有的说法，逐一处理。每一位都写出「看到什么就说明本章对它错，看到什么就说明反之」。

准则污染（Thorndike, 1949; Brogden & Taylor, 1950, 工业与组织心理学）。这是本章最危险的一位近邻，因为它讲的就是「效标里混进了预测源的信

息」，而且比本章早七十多年。分离线在于**它是什么**：准则污染把重合当作要被消除的偏倚，其整个处方是清除它以得到真实的效标关联度；本章主张重合是这类分类得以存在的筛选机制本身，清除它不是得到更真实的估计，而是让被估计的那个东西消失。判定形式：若在把效据完全换成回路外的硬终点之后，证候分层仍能预测差异化的处置反应，则准则污染的诊断对——重合确实只是污染，去掉之后真东西还在；若分层的预测力归零，则本章对——那个分类原本就是被这条回路选出来的。这是两说之间最干净的一刀，且已有部分数据可查。

共同方法偏差（Podsakoff et al., 2003，行为科学方法学）。第二位方法学占位者。它说：同一起来源、同一时点、同一格式的测量会人为抬高相关。分离线在于**重合在哪一层**：共同方法偏差讲的是**测量来源相同**，处方是分离来源（换评价者、换时点、换格式）；本章讲的是**内容上是同一项**，换来源不能解决——由护士评的畏寒和由医师评的畏寒，仍是同一项。判定形式：若在**严格分离测量来源**（不同评价者、不同时点、不同格式）之后，**效应量不下降**，则本章对而**共同方法偏差的诊断不适用**；若效应量随来源分离而**系统性下降**，则那部分确实是**方法偏差**。

替代终点的失效（Fleming & DeMets, 1996，临床流行病学）。极近的一位：抗心律失常药改善了心电指标而增加了死亡率，是这门学问最著名的教训。分离线在于**两个量还是一个量**：替代终点问题是**两个不同的量之间相关性不足**，因此它的处方是验证替代终点与硬终点的关联；本章说的是**判据与效据是同一个量**，不存在相关性问题的，存在的是同一性问题——你无法「验证一个量与它自己的关联」。判定形式：若把**证候积分与硬终点当作两个量做相关分析就能解释全部现象**，则替代终点框架够用；若在控制住二者的相关性之后，**同据率仍独立预测报告效应量**，则本章对。

古德哈特定律与相关的反转模板（Goodhart, 1975；Campbell, 1979；Strathern, 1997，治理与评价研究）。把本章的命题剥掉全部医学名词，只剩「用来判定的量同时被用来验收，于是该量必然达标」。这个纯形式的结构在治理研究里最近的名字就是它：一个指标一旦成为目标，就不再是好指标。必须点名，否则本章会被读成它的一个应用。分离线在于**预测的方向**：古德哈特定律**预测指标达标而实质恶化**——脱钩、粉饰、作弊、挤出；本章**预测指标达标而实质既不恶化也不改善，对回路外是中性的**，因为这里没有人在博弈，条目下降是真的症状改善。判定形式：若**同据率高的体系其硬终点系统性恶化**，则这是古德哈

特，本章多余；若硬终点效应接近零而不恶化，则本章对。这一条极锋利，因为两说预测的符号不同——一个是负，一个是零。

建构性与述行性（MacKenzie, 2006，经济社会学）。它说：模型不是在描述市场，而是在造市场；理论被使用之后世界会朝理论的方向变形。分离线在于**世界变没变**：述行性要求对象本身被改变；本章不要求对象改变——判据与效据重合不需要病人配合，不需要任何人相信什么，两份条目表放在那里，回路就闭合了。判定形式：**若在被判定者完全不知道判定标准内容的情形下，同据率与效应量的关系依然成立，则本章对而述行性不适用**。这是一个可以设计的对照。

克隆选择与自身组织相容性限制（Burnet, 1957; Zinkernagel & Doherty, 1974，免疫学）。这是本章材料的来源，不是竞争者，但必须说清本章比它多说了什么：免疫学用这套机制回答「识别如何可能、耐受如何建立」，它从不追问「判据在不在被选者之外」这个问题，因为在它的问题里这个区分不产生任何后果。本章的贡献不是发现了这个机制，是把它从一个免疫学事实变成一个跨领域的判据。判定形式：**若这套结构只在免疫系统出现，在工程规范与组织例程中找不到对应，则本章的跨领域主张为伪**。

组织例程与演化经济学（Nelson & Winter, 1982，管理学）。分离线在于说不出的原因：例程说主张内容不可完全言明；本章主张内容可以完全言明——七个条目就在标准文本里写着——而**判据不可外置**。判定形式：**若把一套例程或一个证的全部内容完整成文（说得出来），而它仍然无法被回路外的量还原，则本章对而不可言明说不适用于本例**。中医证候标准正是「全部内容都写下来了而仍不可还原」的实例，因此这一条本章当场兑现。

默会知识（Polanyi, 1958, 1966，认识论）。同上，且更普遍地被援引。本章与它的差别只有一句：**默会说把障碍放在表达上，本章把障碍放在参照上**。判定形式：**若把一位名医的全部判断依据形式化为可执行规则（表达障碍被清除），而还原仍然做不成，则本章对**。

循环论证的通行批评（本领域内长期存在的批评意见）。这是与本章最容易被混同的一位，因为结论看上去一样。分离线在于**结论的性质**：通行批评从循环推出「疗效不可信」，是一个否定性的判决；本章从循环推出「这是一个筛选装置，其产物在筛选判据上必然兑现，而在筛选之外没有保证」，是一个带方向的描述，并给出可清点的读数与可操作的处方。判定形式：**若清除回路之后疗效证**

据完全消失，则通行批评对；若清除回路之后回路内的症状改善仍然稳定存在而只是失去了外推许可，则本章对。

证候客观化纲领（沈自尹自一九五九年起的系列研究、陈可冀等的血瘀证研究、证素辨证一路，中医学）。分离线在于**支点**：客观化纲领假定只要找到足够好的指标就能接地；本章指出所有候选指标的确认方式都是与专家判定的一致性，因此它们全部落在回路内。判定形式：**若某一个候选标志物是靠与硬终点的关系、而非靠与专家判定的一致性被确立的，则本章关于「所有确认都在回路内」的论断为伪。**这一条本章邀请对方举反例，且举得出一例即成立。

最近的那一位是谁。是准则污染。它比本章早七十多年，覆盖面更广，工具更成熟，而且它已经说过「判据混进效据会抬高关联」这句话。本章之所以仍不能被它吸收，只有一个理由：**它的整个工具箱建立在「清除污染之后真东西还在」这个假定上，因而它没有位置安放「清除之后那个东西就没了」的情形。**若将来有人证明所有判据与效据的重合都可以在不消灭被判对象的前提下被清除，那么本章应当被撤回，准则污染足够用了。

十四·与本系列各章的分界

本系列近期有五篇是本章的近邻：两篇处理**同一道题**而从不同学科入手，三篇与本章用同样这三门学问。

十四之一·与同题两篇的关系

《随程》（第七章）（同题，从教育学、化学、艺术学入手）说的是：证是一类只在识别程序运行期间成立、并会回写该程序的对象，七十年不收敛是因为每一轮研究在结束前已把自己的对象重划过一次，读数是随程率。**它动的是对象的边界。**

《消参》（第八章）（同题，从物理学、心理学、工程学入手）说的是：证在给定观测结构下结构不可辨识，而可辨识与可控是两个互相独立的性质，读数是消参率。**它动的是对象的内部量。**

本章动的是第三样东西：外部参照点。

三篇的分工可以用一句话说清：一篇说边界在动，一篇说内部分不开，本章说没有回路外的支点。三者互相独立，可以各自成立或各自被推翻。

判决性对照，与《随程》：取一个诊断标准数十年未修订的封闭体系——随程率为零。《随程》预测其机制不发动、指标应当收敛；**本章预测仍不收敛**，因为不可还原来自判据与效据同源，与标准是否修订无关。

判决性对照，与《消参》：取一套判定规则，其内部参数在给定观测结构下全部可唯一确定——消参率为零。《消参》预测其机制不发动；**本章预测仍不可还原**，因为参数即使全部分得开，它们的取值仍然只能用同一批条目去校准，链条还是没有接地。**这一例把本章与那一篇彻底分开**，而且它是可以构造的：设计一套判定规则，参数少、观测通道足够独立，使消参率为零，再看它是否还原得成。

三篇叠加的关系是相乘不是相加：随程率高使跨轮次不可比，消参率高使轮次内部不可辨，同据率高使整条链没有接地点。三者中本章这一条是最难绕开的——因为前两条原则上都可以靠工程手段解决（冻结标准、扩容观测），而**没有回路外支点这件事，靠任何回路内的努力都解决不了**。三篇给出的处方因此有严格的先后：**先加一个回路外的观测（本章），再扩容观测通道（《消参》），最后记录归类变更（《随程》）**。顺序颠倒无意义：在没有接地点的链条上，把参数分得再开、把归类变更记得再全，都拼不出一个外部结论。

十四之二·与同三学科三篇的关系

《叠裕》（《治未病立账学》第三章）（何谓未病）说的是：各单位全合格，而这些合格依据同一份不可分割储备的重复整份计入，读数是叠计比。**分离线：**那一篇处理的是**余量被重复计入**，本章处理的是**判据与效据重合**。**判决性对照：**一个叠计比为一（各单位余量互不共享）而同据率为一的体系，那一篇的机制不发动；本章预测其疗效数据仍然好看而回路外空白。

《功债》（《治未病立账学》第九章）（治未病为何实践最少）说的是：预防不消除扰动，而是把它转成必须一直被供养的残留，读数是续养率。**分离线：**那一篇处理的是**成功之后的代价无出口**，本章处理的是**成功的判据从哪来**。**判决性对照：**一个续养率为零（无残留负荷）的处置，那一篇的机制不发动；本章预测其判据与效据仍可能完全重合。

《漏汰》（《治未病立账学》第六章）（未病如何被识别）说的是识别环节的漏筛与参照脱钩。**分离线：**那一篇讲**该被筛出的没被筛出**，本章讲**筛子的孔是用被筛物自己的材料做的**。两者可以同时发生：一个筛子既漏，又是自制的。**判**

决定性对照：一个漏汰率为零（该筛的全筛出了）的体系，本章预测其同据率可以仍为一——把该筛的都筛出来，并不产生一个回路外的判据。

五篇合起来：一篇记漏了共享余量，一篇记漏了成功的残留，一篇记漏了被漏筛的，一篇记漏了对象的重划，一篇记漏了参数的合并，本章记漏了判据与效据的重合。六种漏记可以同时发生，且叠加时后果相乘。

十五·本章不适用的地方

第一，判据外生的处置。 抗生素、外科复位、疫苗、输血、透析。判定用的量与验收用的量本来就不是一批，本章的机制不发动。

第二，回路外项数不为零的研究。 一项试验若在证候积分之外另附了独立的硬终点，那么它已经在回路外打了一个桩。本章对这类研究的评价不是负面的，恰恰相反——它们正是本章所要的那种研究，本章只要求把这个桩打了没有明确写出来。

第三，关于回路内价值的判断。 本章不对「症状改善值不值得」发表意见。那是一个价值问题，不是一个结构问题，而且答案多半是值得。

第四，同据谱上靠近判据外生一端的那些条目。 同据不是二分。一套判定规则里有的条目从不进入疗效评价（如舌象与脉象在很多方案中只判不评），这些条目是天然的回路外候选，本章预测它们才是最有可能被还原的部分——这是一条正面的、可立即执行的建议：把只判不评的那些条目单独拎出来做客观化，成功率应当高于把全部条目一起做。

第五，本章管不了的一个洞：这条回路是怎么第一次闭合的？在体系形成的早期，判据与效据必定曾经分离过——否则第一批配对无从被筛出。什么时候、因为什么，效据被收进了判据？本章手上的材料不足以回答，姑且交出去。

十六·怎样才算本章错了

十六之一·最强的那个竞争解释

不挑好打的靶子。本章面对的最强竞争解释是这一句：

其实不就是主观指标虚高吗？用病人自述当结局，加上安慰剂效应和期待效应，当然好看。

它能解释本章引用的几乎全部现象：有效率高（主观指标易改善）、跨研究不一致（主观指标不稳）、硬终点上没结果（主观改善本来就不传递）。它不需要任何新概念，而且它是循证医学界对中医疗效数据的标准诊断。

要证明本章不是它的一个花哨说法，必须拎出一个**只有本章的机制成立才会出现、而竞争解释预测不出的变量**，并在控制掉竞争解释之后仍要求它成立。

那个变量是**同据率**。主观指标解释预测：效应量的高低取决于结局指标是否主观，与判据效据是否重合无关——两项都用主观量表的研究，一项用判定条目做结局、一项用另一套独立的主观量表做结局，效应量应当相当。**本章预测这两项会系统性地不同**：用判定条目做结局的那一项显著更高，因为处置正是按这些条目挑出来的。

两个预测在同一份材料上给出可分辨的读数，且都在既有文献里可查。

十六之二·正式的证伪条件

条件一（主条件，机制证伪）：取同一制度环境内已发表的至少六批中医临床研究，全部限定为主观结局，计算每项的同据率与报告效应量。**控制住结局指标的主客观属性、盲法质量、样本量与发表年份之后，若同据率与效应量之间不存在剂量—反应关系，则本章的核心机制为伪**——届时应把疗效数据的形势归因于主观指标虚高，本章多此一举。

条件二（分层预测力，最干净的一刀）：以证候分层为依据施治，用完全回路外的硬终点评价。若分层仍能显著预测差异化的处置反应，**则本章错**——那说明证抓住了某种独立于其判据的东西，重合只是污染，清除后真东西还在。若分层的预测力归零而整方处置的效应不变，**则本章对**。

条件三（符号判决，与古德哈特分开）：同据率高的体系，其硬终点表现应当接近零效应而不恶化。若系统性恶化，**则这是指标失真而非本章所述机制**，本章的解释被更简单的解释取代。

条件四（轮次形态）：本章预测的不是一次先后，是一个循环：**判据先动（外部要求引入硬终点）→ 分类条目在若干年内被改写以对上新判据 → 新条目回写下一版疗效评价标准，被收进积分表 → 下一轮由分类一侧先动，把自己的条目送进判据**。若在六次以上的标准与疗效评价规范修订中，发起处始终在同一侧，**则本章关于发起权轮转的部分为伪**。这可以通过比对历版证候诊断标准与历版疗效评价标准的条目来源来查。

条件五（跨领域）：若「判据与效据重合」这一结构只在中医里出现，而在免疫学实验、工程验收、组织评价里找不到可清点的对应，则本章的跨领域主张为伪，它只是一个领域内的方法学批评。

条件六（来源分离）：若把测量来源严格分离（不同评价者、不同时间、不同格式）之后效应量系统性下降，则那部分是共同方法偏差而非本章所述的内容重合，本章的独有性被削弱。

条件七（只判不评条目）：本章预测只判不评的条目（舌象、脉象等）的客观化成功率高于全部条目一起做。若这两类条目的客观化成功率没有差别，则本章关于「回路外条目更易接地」的推论为伪。这一条可以用现有的证候客观化文献做一次分类统计。

条件八（编码可复现）：若两组独立编码者按第九之一节三条规则对同一批研究方案计算同据率，一致性低于该领域可接受的下限，则同据率不是一个可用的读数，本章的操作部分作废。

十六之三·样本纪律

条件一与条件二涉及跨案例比较，因此必须写死：

- **样本量不少于六个单元，且必须在同一制度环境内取。**可用的单元：同一批国家标准下开展的多项试验、同一学会指南所依据的多项研究、同一省份多家医院的同期研究、同一药物的多项注册试验。
- **严禁用两个国家或两个体系充数。**「中医如此、西医如彼」这类对比在监管、报告习惯、结局选择上处处不同，所谓控制变量在方法上站不住。它可以作启发式例证，**不能充当证伪检验。**
- 若确实只能作跨体系比较，必须在正文中写明这是启发式例证，并另给一个同质多单元的设计顶上。

十六之四·本章可以被分层引用

- **第一层（最强、最易倒）：**同据作为一份有名字的东西，即第④格与第③格有实质差别、且与第①格的差别不在质量而在参照。它被条件二（分层预测力）直接检验。
- **第二层（分析工具）：**判据外生与否、有无持续筛选两根轴与四格。即便第一层被推翻，这张表仍可用于给任何一个评价体系定位。

·第三层（最稳、最便宜）：一条不依赖前两层的经验主张——中医临床研究中判定条目与疗效条目的重合项数从未被系统统计过，而它只需要两份文本就能算出来。这一条不要求接受本章的任何理论，它只指出一件没人做过、成本近乎为零、而结果会直接影响如何解读三十年疗效数据的事。

说明本章可以被分层引用，是诚实，也是保护。

十六之五·若本章被证伪，会学到什么

若条件二失败——用硬终点评价时证候分层仍有预测力——那将是本章所能带来的最好的消息，而且它会把资源引向一个明确的方向：那意味着证确实抓住了某种独立于其判据的东西，而找到它的正确做法是从此以硬终点为筛选判据，而不是继续用证候积分。这个结论比本章原来的主张更有价值。

十七·这个判断被采用之后会怎样反噬

反噬一：要求报告同据率，最省事的达标办法是改写条目措辞。把疗效量表里的「畏寒」改成「怕冷程度」，把「便溏」改成「大便性状」，两份表在字面上就不重合了，同据率降到零，而实际测的还是同一件事。规则一之所以必须要求两名独立编码者判定「同一项」并报告一致性，就是为了挡这一手；但编码本身也可以被优化。

反噬二：要求引入回路外终点，最省事的达标办法是加一个几乎不会变化的硬终点。在一项六周疗程的试验里加一个「全因死亡率」，几乎必然是零对零，形式上满足了「回路外项数不为零」，实质上什么信息也没有。第二字段之所以要求分别报告硬终点项数与其可变化性，就是为了挡这一手。

反噬三：把这个判断讲出来，最省事的接收方式是否定病人的好转。「回路内有效」这个词一旦流传，很容易被简化成「假的有效」。第九之四节与第十二之一节两处正是为了挡住这个方向，但削减写在论文里，口号传得比论文快。本章对此的处理是把这条写进伦理边界的最前面，而不是留在讨论部分。

出口在哪里？我没有找到。任何要求把判据与效据分开的规则，都会遇到一个真实的困难：对于以主观症状为主要负担的疾病，症状本身就是最重要的结局，硬要另找一个回路外终点，往往找不到比症状更有意义的。三条路各有各的坏：不分开，则数据不承载外部结论；强行分开，则可能把最有价值的结局挤出

去；加形式化的硬终点，则得到一堆零对零。本章能做的只有把这三条写在这里。

伦理边界另有三条，且这三条排在本章全部结论之前：

其一，**回路内有效是真实的价值，不是假象。**畏寒减轻、睡眠改善、食欲恢复，这些就是病人的生活。本章的任何部分都不得被用作削减对症状改善的支付、供给或研究投入的依据。

其二，**本章不主张取消证候疗效评价，只主张把它与回路外结局分栏报告。**取消它会同时取消这套分类赖以被筛选的机制，其后果比现状更坏。

其三，**在恶性、进展性、有明确有效外部疗法的疾病中援引本章的，必须同时援引第十二之一节与本节前两条。**只引「回路内有效不传递」而不引「传递是可能的、且症状价值是真实的」，是误用；而在这类疾病中以回路内数据替代回路外证据作出治疗决策，是本章明确反对的。

十八·本章自己落在哪一格

一条关于「判据与效据不能同源」的判断，若不能用在自己身上，它就是自封的。

本章提出的同据率，按本章自己的两根轴判定，落在哪一格？

本章用来判定一项研究「有问题」的依据，是同据率高；而本章用来验证自己这个判断成立的依据，是同据率与效应量的关系。**两者用的是同一个量。**若有人问：你凭什么说同据率是个问题？本章的回答是：因为同据率高的研究效应量更高。而效应量为什么算问题？因为它是同据率抬起来的。

这是一个真实的循环，本章必须认下来。本章的第②条与第③条证伪条件正是为了给自己打回路外的桩：条件二用的是硬终点下的分层预测力，条件三用的是硬终点的符号——两者都不依赖同据率与效应量的关系。**本章的成败，压在这两条上，而不压在本章自己最喜欢的那条剂量—反应关系上。**若只有条件一成立而条件二、条件三不成立，本章就是第④格里的又一个样本。

后果一：本章不能靠被引用来证明自己。引用本章的人多半是已经接受「循环有问题」这个前提的人，他们的引用不构成回路外的证据。

后果二：本章给自己设了一条时限。若到二〇三六年，条件二从未被任何独立团队执行过一次——这件事所需的只是一批以硬终点为结局、以证分层的研究

——那么本章与它所批评的对象处在同一格：一个从未被回路外检验过的判断。届时本章的第一层应当被撤回，只保留第二层的分类工具与第三层的经验主张。

后果三：本章所处的位置正是它所描述的那一格。学术评价也是一套判据与效据高度重合的体系：判定一篇文章有价值用的是新颖、严谨、可检验，验证它确实有价值用的也是这几条，而不是它此后在世界上引起了什么。本章若被接受，多半不是因为它对，而是因为它在这套回路内兑现；本章若被拒绝，也未必是因为它错。这句话对本章不利，但按本章自己的判断，它必定成立。

十九·结论，与一条写死日期的赌注

七十年的还原研究做不成，不是因为对象太复杂、样本太小、技术太粗。是因为还原需要一个支点——一个独立于被还原对象的参照——而在一条判据与效据同源的回路里，这个支点不存在。每一次「找到了标志物」，找到的都是与判定条目共变的量；而判定条目的正当性来自它在疗效上兑现，疗效又用判定条目来测。整条链上没有一处接地。

它有效，也来自同一条回路。数亿诊次的筛选把不兑现的配对全部逐出，剩下的必然兑现。这份兑现是真的，只是它的范围止于筛选所用的那份判据。回路内有效与回路外有效，是两件事，而现行的话语里没有把它们分开的词。

承认这一点的代价，不是承认中医无效，而是承认三十年的疗效数据回答的不是它们被用来回答的那个问题。这笔账的规模，取决于同据率有多高；而这个数只需要两份文本就能算出来。

最省事的第一步不需要任何理论上的让步：取一批已发表的中医临床研究方案，把诊断标准条目表与疗效评价条目表并排放，数一数重合项，报出同据率与回路外项数。不需要新的病人，不需要新的经费，不需要同意本章的任何一条判断。三十年里，这件事没有被系统地做过一次。

给出一条写死日期的赌注：

到二〇三四年十二月三十一日，至少有一份中医临床研究的报告规范、注册要求或指南制定手册，明文要求把「证候诊断所依据的条目」与「证候疗效评价所依据的条目」分别列出，并报告二者的重合项数。

什么不算命中，一并写死：

1. 该要求只出现在方法学论文、社论或综述中，未进入任何规范、注册要求或制定手册，不算命中。
2. 只要求「鼓励采用客观指标」或「建议增加硬终点」，而不要求把两份条目表分别列出并报告重合项数，不算命中——那是一个方向，不是一个字段。
3. 只要求实施盲法或安慰剂对照，不算命中——那处理的是另一个偏倚，与判据效据是否同源无关。
4. 只在个别期刊的投稿须知中出现，而未进入任何行业层面的规范或注册要求，不算命中。
5. 该条款由本章作者或其直接合作方参与推动写入的，不算命中——赌的是这件事会自己发生。

若到期未兑现，本章关于「这个字段的缺失是结构性的」这一判断得到反面证据：那说明它不是一个被制度性遮蔽的位置，只是一件还没轮到去做的事，本章对其结构性的估计过高了。

注释

〔一〕本章所称「判据」，指一次分类判定所实际依据的、可被单独记录取值的观测项集合；「效据」指判定该次处置有无效果所实际依据的同类集合。二者均以研究方案或标准文本中的明文为准，不以研究者事后的说明为准。

〔二〕本章所称「回路外」，指从未被用于该体系内部筛选的判据。它不等于「客观」——一个从未被用作筛选依据的主观量表同样是回路外的；也不等于「硬终点」——硬终点只是回路外判据中最不易被同化的一类。

〔三〕第九节的同据率与准则污染在算法上有交集，但报告要求不同：准则污染的处理是清除后重估，本章要求把清除前的原始重合数报出来。**两者的差别不在算法，在于报不报。**

〔四〕第十三节对各既有说法的表述力求公允，未作简化；若有失当，责任在本章。

〔五〕第十六节条件二所需的研究设计（以证分层、以硬终点评价）目前极为稀少，这是本章没有自己完成该项检验的原因，也是本章把它列为最关键的一条证伪条件而非已完成证据的原因。**这条检验的稀少本身，就是本章所描述的形势的一个例证；但本章不把它当作证据，因为那正是循环。**

第四编 结算

九栏合起来是什么

第十章 同一组学科被问了三遍

一·这一编要做的事

前九章各占一格，各自与本领域的近邻划过界，各自交出一个读数与一句判据。九章合起来，还有一件每一章都不必回答、而一本书必须回答的事：**这九样是不是同一样东西？**

如果是，那么九章就是九个换了说法的同义反复，本书应当变薄到只剩一章；如果不是，那么必须说清它们凭什么不是，而“看上去不像”不构成理由。第四编的五章分头处理这件事：本章处理设计层面（九章是怎么被生成出来的，这个生成方式本身能不能排除“九样是同一样”），第十一章处理测量层面（九个读数彼此是不是独立的），第十二章处理执行层面（九句判据合起来能不能变成一份可以拿去用的东西），第十三章处理合成层面（九栏之间是相乘还是相加），第十四章处理后果层面（把这九栏立起来会引出什么新麻烦）。

本章先交代设计。

二·九章排在一张三乘三的表上

九章不是九次各自独立的发现。它们是三组学科与三个问题交叉之后的九格：

	A 化学·教育测量·文物与艺术品保存	B 物理／计量·心理·工程	C 细胞／免疫·工程·管理
What 一个证是什么	一 让层	二 吞差	三 配格
How 它凭什么被复核	四 并身	五 摊差	六 填判
Why 为什么七十年还原不成而临床有效	七 随程	八 消参	九 同据

导论第六节已经说过这张表怎么读，此处只补两笔它不能给出的东西。

它不是一个排除法。横读与竖读合起来，排除的是两种最方便的还原——“九样是那几门学科的产物”与“九样是问题措辞逼出来的”。排除这两种之后，剩下的解释不止一种。至少还有三种没有被排除：

其一，同一个人同一套模板。九章由同一位作者按同一道工序写成——三条互不相容的答法、共同假定、从其中一家自己的材料里找推翻它的那一件、命名、比率、判据。这道工序本身可能就是一台“生产被吸收的量”的机器：把它对准任何题目，都会产出一样形状相同的东西。这一条本书无法排除。

其二，学科与材料经过挑选。三组九门学科不是随机抽的，是作者选的；每一章“推翻共同假定的那件材料”也是从该学科的现成成果里挑出来的。挑选的过程留下的痕迹，在成稿里看不见——读者看到的是“化学自己的保护基推翻了三家共同假定”，看不到的是有多少条化学材料被试过而没用上。

其三，后验命名。九个名字都是在看见那样东西之后起的。事后给一个已经被指认出来的空白起名，几乎总能成功；成功本身不证明那个空白是实在的。

因此本章对这张表的表述必须收窄为：三乘三的设计对“被吸收是这类判定本身的性质”这一判断构成支持，而不是排除法给出的唯一解。唯一的出路写在本章第六节。

三·三处继承下来的盲区

同一组学科被用了三次，那一组的盲区就被继承了三次。三组各有一处，必须逐条指认，因为同一组内的三章会因为同一个理由一起出错，而不是依次出错。

三之一·A组：它们都拥有一份“处置记录与对象记录分立”的档案传统

化学的操作规程与分析报告是两份文件；教育测量的施测记录与作答记录是两份文件；文物与艺术品保存的现状病害报告与修复处理报告是两份文件，而且后一对的分立是写进宪章的硬规定。

正因为这三家都有这个分立，作者才能从它们那里看出中医没有这个分立。第一章的让层、第四章的并身、第七章的随程，处方在形式上都是同一件事：把处置那一侧从对象那一侧分出来单独记。

盲区就在这里。三家没有一家处理过“两份记录从一开始就不可能分立”的体系。中医的病历恰恰是这种：四诊所见与据以处置的判断写在同一段文字里，且后者决定了前者要采哪几项——一位打算温阳的医生问的问题与一位打算清热的医生问的不一样，采集本身已经被处置意图挑选过。这不是记录习惯不好，是这种判定的采集与处置在时间上不可分。

后果：A组三章的处方，在“采集被处置意图挑选”的体系里可能原理上执行不了，而不只是执行起来困难。三章都没有正面处理这一情形，因为它们的三个来源都没有这一情形。

三之二·B组：它们的三个分母都要求先能数清“本来有多少”

物理计量有溯源链接到定义常数，心理测量有真分数模型，工程有基准与包线。第一本书已经指认过这一组共同默认“存在一把独立于被测者的刻度”。在本书里，这个盲区换了一个更具体、也更麻烦的形状——它落在分母上。

第二章的吞差率是 $(k - m)/k$ ，其中 k 是“把首末两次记录直接并排会判出的改变次数”；第五章的存异率的分母是“这次复核中出现过的分歧条目数”；第八章的消参率的分母是“原始参数数”。三个分母都不是从现成记录里数出来的，它们都要求先能说出“本来有多少”——而“本来有多少”只有在存在一把外部刻度时才有答案。

三章各自都意识到了自己那一处（第二章说 k 须由一次盲判给出，第五章说分歧条目须在过程中埋点，第八章说参数数须先写下一个模型结构），但没有一章知道另外两章有同一个毛病，因为每一章只占一格。

后果：B组三章会在同一类领域里一起失效——凡是没有独立刻度、且过程不可埋点的场合，三个读数同时算不出来。这不是精度问题，是分母不存在。

三之三·C组：它们全都是“选择—保留”式的，因而看不见从未进入筛选的那一类

细胞学讲克隆选择，工程学的规范是用事故写出来的，管理学讲例程的选择与保留。三家共享一个背景信念：长期运行的淘汰机制会把不好的清掉，因而留下来的东西携带着信息。

第九章正是靠免疫学自己推翻了“选择的判据在被选择者之外”，这是本书最漂亮的一次自毁材料。但推翻的是判据的位置，没有推翻这个背景信念本身。

盲区因此是：第三、六、九三章都是在筛选装置内部说话。配格问的是“这一个能不能替换那一个”，填判问的是“这一个属不属于这一型”，同据问的是“判定与效果评价用的是不是同一批条目”——三问都预设被谈论的东西已经在装置里了。三章一起看不见从未进入筛选的那一类：既没有被留下也没有被淘汰，只是从来没有被判定过。

这一类在中医里有一个具体的名字：那些从未被任何一位医生辨过证的人。他们不在任何一份病历上，因而不在于任何一个读数的分母里。

后果：C组三章的读数全部是有偏的，且偏的方向一致——它们只在已被判定的对象上有定义，而“哪些人会被判定”本身是被同一套装置决定的。

四·本书自己的叠计比

第一本书的第三章给过一个读数：叠计比——若干个单位各自声明合格，而这些合格依据的是同一份不可分割的储备被重复整份计入，则这份储备被计了几次，叠计比就是几。那一章还说，这个读数应当先用在提出它的人自己身上。

照办。工程学在本书九章里出现了五次：第三章（公差带与互换性制造）、第五章（安全论证与开放项关闭）、第六章（型式试验与使用包线）、第八章（系统辨识与激励设计）、第九章（失效反选规范）。

五处取用的是工程学五个互不重叠的分支，这是本书的一处辩护：不是同一条结论被引用五次，是五条不同的结论各被引用一次。因此**若某一条工程学结论被推翻，只有那一章动摇，不会五章同时动摇。**

但这个辩护挡不住另一层。五章共享的不是某一条结论，是一个更底层的设定：**工程学是一门“把判定条件显式写下来”的学问，因而它可以充当其余学科的对照家。**型式试验把使用包线写在合格证上，安全论证把开放项列成清单，系统辨识把可辨识性写成一个可以在取数据之前判定的结构性质——本书五次拿工程学当镜子，靠的都是这一条。

若这一条被推翻——例如被证明工程实践中真正起作用的判定条件绝大部分同样不成文，写在合格证上的那些只是仪式——那么五章会同时失去它们的对照面。这不是五章各自的论证失效，是它们共用的那面镜子碎了。

按第一本书那个读数的定义，**本书的叠计比至少是五。**而这五次取用没有在任何一章里被加总过——每一章只说“本章借用了工程学的某某分支”，没有一章说“本书已经借了第五次”。这正是那一章所指认的形状，出现在指认它的人自己的书里。

五·与第一本书的一处交叉核对

两本书用的是同一张学科表。这带来一件第一本书做不到的事，也带来一笔必须算清的账。

做得到的事：跨题域的第三向。第一本书的三乘三证明了，在“未病”这个题域里，缺栏不是学科的产物也不是问法的产物。本书的三乘三在“辨证”这个题域里得到同一个形状。**两本合起来，同一张学科表被两个不相干的题域各问了一遍，交出的是形状相同而内容毫无重叠的两批量。**这一向是任何一本单独都给不出的：它排除的是“这张学科表本身有偏”——若这九门学科天生倾向于产出“被漏掉的量”，那么两个题域产出的应当彼此相似，而实际上没有一样重合。

必须算清的账：盲区被继承了六次，不是三次。第一本书第十章已经指认过三处继承盲区（A组三家各自自带一个照料者角色、B组三家默认存在独立刻度、C组三家默认单元可替换）。本章指认的三处（A组默认两份记录可分立、B组三个分母要求外部刻度、C组三章在筛选装置内部说话）与那三处**不是同一批**，但它们来自同一组学科。

这意味着：**A组三家一共给这两本书带进了两处盲区，B组与C组同理。**两本书的读者若把两本合读，会同时继承六处，而其中B组那两处（默认独立刻度／分母要求外部刻度）在形式上是同一件事的两面——**这一处是六处里唯一重复的，也因此是最该被优先补掉的一处。**

六·本书自己无法检验的三条，以及各自该请哪一门学科

第一条：九个量在没有独立刻度的领域还成不成立。本书无法自查，因为九章的三组学科里有一组正是靠刻度吃饭的。**该请的学科：历史语言学的比较重建法。**它是少数在完全没有外部刻度的条件下重建变化的成熟传统——链移研究正是第二章借过的那一处，但第二章只借了“相邻不可分辨会累积成整体漂移”这一条，没有借它“如何在无刻度下重建”的那一整套。

第二条：处置记录与对象记录从一开始不可分的体系，A组三章的处方还执行不执行得了。**该请的学科：司法的证据保管链与讯问笔录制度。**讯问笔录恰恰是这样一份文件——提问本身塑造了回答，而制度对此的应对不是把两者分开（分不开），是把提问一并录下来。这是本书三组学科都没有的第四种路子。

第三条：从未进入筛选的那一类。该请的学科：保险的核保与拒保档案。保险业是少数系统性登记“没做成的那一笔”的行业——拒保记录、除外责任、未成交询价都要留档。C组三家没有一家有这个传统。

三条各自指名了下一门该请的学科，而不是笼统地说“有待进一步研究”。这是本书能做到的最诚实的一步：**指出自己的三个洞在哪里，并说出该由谁来补。**

七·本章的位置

本章不新增读数，不新增判据，也不新增可被证伪的预测。它做的是一件辅助工作：把前九章的生成方式摊开，指出这个生成方式能支持什么、不能支持什么，以及它在哪三处系统性地看不见东西。

按本书自己的规矩，一章没有读数也没有判据是一次违规。这次违规是自觉的，理由是：**一本书对自己的设计做交代，若也要配一个比率，那个比率只会是装饰。**但违规就是违规，记在这里。

第十一章 九个读数放在同一张表上

一·为什么必须并表

九章各给了一个读数。九个读数分开看，每一个都清楚；放在一起，立刻有三个问题浮上来：它们量的是不是同一样东西？它们能不能加起来？谁能真的把它们采齐？

前两问关乎本书成不成立，第三问关乎本书有没有用。本章逐个处理。

二·九个读数的分子、分母与型

章	读数	分子	分母	分母类型	型
一	让层比	本次处置未予针对的层数	本次判定认定成立的层数	项数	C
二	吞差率	$k - m$ （首末直接比较判出的改变数，减去记录里登记过的改变数）	k	次数	M
三	配格率	为使这一次配得上而做的调整量	该次处置的总量	项数	D
四	留底率	取自首次处方之前已固定、或结构上不受其影响的证据项	复核所依据的全部证据项	项数	D
五	存异率	归档文件中仍可检索到技术内容的分歧条目	这次复核中出现过的分歧条目	事件数	M
六	弃判率	记为“不适用／无从判定”且带下游处置路径的次数	一段时期的全部判定输出次数	次数	D
七	随程率	由程序或标准变更引起的对象归属变更次数	最近 N 次对象归属变更	次数	C

		数			
八	消参率	原始参数数－可唯一确定的参数组合数	原始参数数	结构量	M
九	同据率	判定依据项集合与效果判定依据项集合的交集	判定依据项集合	项数	D

型沿用第一本书的三分：**D** 直接审计（翻现成文件即可数出）；**C** 重构编码（须先按规则把现成记录重新编码）；**M** 模型依赖（须先写下一个模型、做一次实验或在过程中埋点）。

九个里 **D 型四个、C 型两个、M 型三个**。这个分布比第一本书好——那一本九个里只有一个 D 型。原因不难看出：本书的对象是判定这个动作本身，而动作留下的痕迹（处方药味、证据项日期、判定输出次数、条目表交集）比对象的状态更容易被直接清点。**这是本书相对第一本的一处实质优势，应当被利用：先跑四个 D 型的，再谈别的。**

D 直接审计 C 重构编码 M 模型依赖 九个里 D 型四个，先跑这四个

图五 九个读数的型与成本档

三·一个不可得的分母，必须单列

第五章的存异率有一个别的读数都没有的毛病：**它的分母原则上事后不可得。**

其余八个读数的分母都能从现成材料里数出来或算出来——层数可以从病历重构，药味可以数，证据项有日期，判定输出有条数，归属变更有版本记录，参数数由模型结构给定，条目表可以并排放。**唯独“这次复核中出现过的分歧条目数”不能。**理由正是第五章要论证的那件事：分歧在结论形成的那一刻被消灭了，被消灭意味着它不留痕，不留痕意味着你事后无法知道它有过几条。

这不是一个技术困难，是这个读数与它所要测的东西之间的一种内在关系：**存异率所测的正是“分歧留没留痕”，而分母的可得性恰恰依赖于分歧留了痕。**分母能数清的时候，答案已经接近一了。

因此存异率的正确用法只有两种，第五章给了第一种，本章补第二种：

其一，前瞻埋点。在复核过程中先设一个“未采纳意见”栏，此后每次复核都填，分母由此产生。这是第五章的处方，它同时是这个读数的采集条件——**处方与量具是同一类东西，这在九个读数里是唯一一例。**

其二，从下游倒推一个下界。取一批“复核结论后来被推翻”的病例，看当初的归档文件里能不能查出有人提过异议。查得出的比例是存异率的一个下界，且这个下界不需要任何前瞻设计。它测不到真值，但它足以判断一个体系是不是在零附近。

本章建议第二种先做——因为它属于 D 型，一个下午可以做完；第一种要等两年才有数。

四·三条不做加权总分的原因

一个自然的念头是把九个读数加权合成一个“结算指数”，一个数说清一个机构的好坏。本章明确拒绝，理由三条，任何一条单独成立即足够。

其一，分母不同类。表里的分母分属四类：项数、次数、事件数、结构量。把“这张方里超出基础方的药味占比”与“过去一年被填过几次无从判定”加权平均，得到的数没有任何一个可以被解释的量纲。这不是精细与否的问题——四类分母之间不存在换算。

其二，其中一个的分母不可得。只要存异率在表里，总分就永远算不出一个可复现的值：任何一个机构都可以通过不埋点而使这一项无从计算，而“无从计算”在合成时会被当成“缺失值”处理掉——**结果是越不留痕的机构，这一项越不影响它的总分。**这是一个把处罚变成豁免的设计。

其三，也是最要紧的一条：在测出相关之前做总分，正是本书第三章指认的那个错。第三章说的是“归为同一类不蕴含在使用处可以互相替换”。把九个读数加权求和，做的正是把它们当作可互换的度量对待——每一个都被换算成同一个尺度上的一份贡献。**本书若做总分，就是在自己的书里犯自己第三章指认的那件事。**

因此本书的立场是：**九个读数并排报告，不合成。**需要一个概括时，报告的是“九项中有几项落在警戒区”这样一个计数，而不是一个加权分。计数不预设可互换性。

五·五对相关的预注册

九个读数两两共三十六对。本章不对三十六对全部下注，只对最可能相关的五对做预注册——写在见到数据之前，且写明若结果如何则本书须做什么。

第一对：让层比 × 配格率。两者同属“处置只覆盖了对象的一部分”这一族，是九对里最可能重复的一对。**预测 0.3 至 0.6。**理由：让层比数的是被认定成立而未动的层，配格率数的是为配上这一个人而做的调整——前者是没做的，后者是多做的，方向相反但同源。**若超过 0.8，两章必须并章，保留配格率——**因为它是 D 型，可从处方直接清点，而让层比要从病历重构。

第二对：留底率 × 存异率。第五章已经指出两者都低时可回溯性接近相乘而非相加。**预测低相关（0.3 以下），**但两者的乘积对“复核结论被推翻后能否查出当初争议”具有独立解释力。这一条比单纯的相关系数更值得测：**若乘积项没有独立解释力，第五章与第四章的那段叠加论证失效，**尽管两章各自仍可成立。

第三对：弃判率 × 同据率。预测负相关。理由：同据率高意味着判据与效据同源，筛选回路闭合得好，体系内部很少出现“这一例我判不了”的处境——因为总有一个条目可以填。**若实测为正相关或零相关，第六章与第九章之间缺一条本书没有写出的中介机制，**须补写。

第四对：随程率 × 消参率。预测正相关，**0.4 至 0.7。**理由写在第八章：规范化在结构上就是激励削减。同一轮规范化既重划对象（抬高随程率）又削减激励（抬高消参率）。**若超过 0.8，两章必须并章，保留消参率——**它是唯一一个不需要采集任何新数据、纯由模型结构算出的读数，成本最低。

第五对：吞差率 × 留底率。预测近零。一个测跨次漂移，一个测单次复核的证据来源，两者在机制上无接触。**若实测显著相关，说明存在一个本书没有识别的共同因子——**最可能的候选是记录制度本身（一份写得详尽的病历同时抬高两者），这将要求在两章各加一个控制变量。

统一规则，写在这里以免日后找理由：任何一对相关超过 0.8，必须并章，不得以“侧重不同”“视角互补”为由保留两个名字。这条规则的效力高于本书任何一章的自我辩护。

同时收回一句话。上述五个数值区间是预测，不是阈值；0.8 这条线是并章的触发条件，不是“低于 0.8 即可宣告独立”的通行证。是否并章还要同时看六

件事：定义是否重叠、控制测量误差后潜变量相关有多高、各自是否对结果有独立预测力、在不同场所方向是否一致、并章后会不会丢掉一个可执行的判据、以及并章后剩下那个读数的采集成本。只看一个相关系数就并章或不并章，都是偷懒。

六·九个读数在任何单一场所都采不齐

这是本章最要紧、也最不利于本书的一节。

临床病历这个场所能采到五个：让层比（重构处方针对的层与判定认定成立的层）、配格率（数药味）、留底率（看证据项日期）、弃判率（数判定输出）、同据率（并排放两份条目表）。采不到的四个是：吞差率（须另做一次首末盲判）、存异率（须前瞻埋点）、随程率（须有跨版本的标准修订记录）、消参率（须先写下一个模型结构）。

研究文献这个场所能采到三个：随程率（标准修订史）、消参率（模型结构）、同据率（方案书里的诊断标准与疗效评价条目表）。采不到的是全部与单次判定过程有关的读数。

复核过程这个场所能采到存异率与吞差率，但必须事先设计，事后一条也补不上。

三处的交集只有同据率一个。这意味着：

九个读数的完整相关矩阵，在任何单一场所都做不出来。

只能分场所测子矩阵，再设法拼接。而分场所拼接会把**场所差异**混进相关估计——两个读数若在临床上测、另两个在文献上测，它们之间的相关里必然含着“临床与文献这两批材料本身的差别”，而这个差别与两个读数各自要测的东西无关。

这一条直接削弱第五节的全部预测。第五节的五对里，只有第一对（让层比 × 配格率）与第三对（弃判率 × 同据率）可以在同一场所测得；第二对（留底率 × 存异率）跨临床与过程埋点，第四对（随程率 × 消参率）跨文献与建模，第五对（吞差率 × 留底率）跨过程与临床。**也就是说，五对预测里有三对，其实证结果原则上无法与场所差异分开。**

本书没有解决这个问题，只能给出一条不彻底的对策：**在同一批对象上同时开两个场所**——选一家有完整病历、有本院标准修订史、且愿意在复核流程里加

一个字段的机构，跟满两年。这样九个读数中的八个可以在同一批对象上采齐（吞差率仍须一次专门的盲判）。**本书不知道有哪一家机构在做这件事，也没有去找过。**

⇒ 完整相关矩阵在任何单一场所都做不出来；分场所拼接会混入场所差异

图六 九个读数在三个场所的可采性

七·本书自己的九个数

九章各自在倒数第三节算过自己的读数。汇总在此，因为它们合起来构成本书对自己的一次结算：

- **本书的叠计比至少是五**（工程学被取用五次，未加总；第十章第四节）。
- **本书的让层比不为零**：九章中的每一章都认定了若干个相邻问题成立而本次不动——第一章认定“证与病的关系”成立而不处理，第七章认定“识别程序为什么会被修订”成立而不处理。这些被让开的层，本书同样没有为它们设字段。
- **本书的留底率偏低**：九章的三组学科与九个题目，是在写作过程中被反复挑选与调整的，而挑选之前的候选清单没有留底。读者拿到的是终稿，看不到有多少学科组合被试过。
- **本书的存异率为零**：写作过程中被否掉的判断、被削弱的主张、被换掉的读数定义，在成书里只留下“版本”一行的简述（如第三章的比率定义曾被改过、第二章的比率定义由不可清点的形式改为可清点的形式），**没有一处保留了被否掉的那一版的技术内容**。这是本书自己指认的那件事发生在本书自己身上，且这一次没有借口——加一个附录就能解决。
- **本书的弃判率为零**：九章没有任何一处写下“这一题本书判不了”。第十章第六节列了三条无法自检的判断，但那是“无法自检”，不是“无从判定”，两者不同。
- **本书的同据率接近一**：九章用来提出主张的材料，与用来支持主张的材料，是同一批——三组学科既是提出问题的来源，也是兑现预测的场所。第九章说这种结构使还原没有支点。**本书自己正落在这一格里。**

最后一条不是修辞式的自谦。它有一个具体后果：**本书无法用自己的方法证明自己**，任何以九章材料为依据的自我验证都在回路内。这与第十章第六节那条“须由另一组不知情的研究者重跑”是同一件事的两种说法。

八·三档成本

给要试的人一个次序，按代价从低到高。

甲档：一个下午，翻现成文件即可。 同据率（把诊断标准的条目表与疗效评价的条目表并排放，数交集）、弃判率（数一段时期的判定输出里有几次记为无从判定）、留底率（看一份复核所依据的证据项各自的日期）。**这三项同时可以证否本书第三层的实践主张：**若这三个数在多数机构里已经处在合理区间，本书的处方就没有必要。

乙档：一周到一月，须重新编码。 让层比、配格率、随程率。前两项要从病历与处方重构，第三项要翻标准修订史。

丙档：须设计、埋点或建模。 吞差率（一次首末盲判）、消参率（先写下模型结构再做结构可辨识性分析）、存异率（前瞻埋点，两年起）。

本书一次也没有跑过甲档。 这是导论里已经认过的第一笔账，在这里再认一次——因为它就在这一节的正上方，而且只要一个下午。

第十二章 九句判据合成一份问卷

一·九句原文

九章各自交出一句不含情态词的问话。此处原样抄录，未作任何改写——改写会使它们失去唯一的价值。

第一章·让层。

这一次的处置针对的项，与这一次判定认定成立的项，数目是否相同？不同的那些记在哪个字段？

第二章·吞差。

把第一次记录与最近一次记录直接并排，不看中间，会不会判出同一个证？记录里登记过几次证型改变？

第三章·配格。

这一张方里，有几味在基础方之外？这几味针对的是这个人的哪一个特征，记在哪个字段？同一时期辨出同一个证的另一位病人，能不能用这一张方？

第四章·并身。

这一次复核所依据的记录里，有哪几条是在上一次开方之前就已经写下的？

第五章·摊差。

这次复核归档的文件里，“未采纳意见”有没有一个字段？上一次填的是什么？

第六章·填判。

上一次复核的结论表上，“无从判定”有没有一个可填的格？过去一年被填过几次？

第七章·随程。

最近 N 次分类变更中，有几次的当日记录里没有出现新的观察项？

第八章·消参。

在这一次判定所依据的全部观测项里，有多少对在过去 N 次记录中始终同时出现、同时缺失？

第九章·同据。

这一次判定所用的观测项，与判定这一次有没有效所用的观测项，有几项是同一项？

二·三条共同性质

九句是九位作者式的产物吗？不是——它们是同一道工序在九个位置上的产物，因而共享三条性质，而这三条正是它们能被合成一份问卷的理由。

其一，没有一句带情态词。九句里没有“应当”“可能”“是否合理”“有没有必要”。全部是数数（有几项、几味、几次、几对）或查有无（有没有一个字段、有没有一个格）。这不是文风，是设计：**带情态词的问题只能被回答者的判断回答，不含情态词的问题只能被记录回答。**

其二，答案必须去查，不能靠回想。九句没有一句可以在受访者的印象里得到答案。“上一次填的是什么”要去调出上一次的归档件；“过去一年被填过几次”要去数。这使得问卷的效力不依赖受访者的诚实——**一个想给出好看答案的人，唯一的办法是去改记录，而改记录会留下痕迹。**

其三，问的是这一次判定，不是这个对象。九句没有一句在问病人。它们全部在问那一次判定：它针对了几项、它吸收了什么、它的依据来自哪里、它有没有拒绝的余地。**这一条决定了这份问卷的伦理位置——它诊断的是制度，不是人，也不是病。**

三·每句展开三问：二十七问

九句是判据，不是问卷。判据要变成问卷，每一句须拆成三问：**有没有这个字段（存在）／这一次填了什么（内容）／过去一段时期填过几次（频次）**。三问不可省，理由是它们各自能证否不同的东西：存在之间证否“这一栏根本不在”；内容之间证否“栏在而这一次没用”；频次之间证否“栏在、这一次用了，但它只是被偶然用了一次”。

九句 × 三问 = 二十七问。全文列在附录二，此处只举第五章那一句作示范：

存在：本机构的复核归档模板里，是否存在一个专供记录未被采纳意见的字段？（是／否／有自由文本区但不与结论同屏）

内容：最近一次复核的归档件里，该字段填写的内容是什么？（原文抄录／为空／字段不存在）

频次：过去十二个月的复核归档件中，该字段非空的份数占比是多少？（数字／无法统计）

三个选项里的第三项在每一问里都不同，且都不是“其他”。它们各自对应一种本书预测会大量出现的中间状态：有自由文本区但不与结论同屏（第六章已指认这等于不存在）、字段不存在（与“为空”须分开记）、无法统计（本身就是一个答案）。

四·三条编码规则

规则一：字段必须在被判定者的那一份记录上。记在科室台账、质控报表、研究数据库里的，不算。理由是本书九章处理的都是“这一次判定”，而下次要用到它的人只会看到那一份记录。

规则二：可选填不算。一个字段若在系统里可以跳过而不影响提交，记为“有而不构成约束”，在统计时并入“无”。这一条会显著拉低所有机构的得分，本书接受这个后果——第十四章会说明为什么不能放宽。

规则三：只记有无，不记内容质量。一个填了“无异议”的未采纳意见栏，与一个填了三行技术理由的，在本问卷里同样记为非空。理由不是本书觉得两者一样，是一旦开始评判内容质量，问卷就重新需要评判者，而九句问话的全部价值在于不需要评判者。内容质量属于另一份工具，本书不提供。

规则三有一个代价，必须写明：**它使问卷可以被最省事地满足**——所有字段设立、一律填“无”。第十四章第二节把这一条列为本书预言的最可能结局。

五·一次设想走查

以下不是真实审计。它是一次纸上走查，目的是让二十七问在一份具体材料上跑一遍，看会出什么。本书没有做过真实走查，这一点在导论与第十一章都已认过。

设想一份档案：一位六十八岁患者，四个专科随访，中医科每月一诊，连续两年，病历完整规范，字迹清楚，四诊记录逐次齐备，证型判定连续，处方逐次可查。**这是一份优秀病历**——在任何一次质控检查里都会被被评为优秀。

二十七问跑完的结果：

九栏中八栏为“无”。处方针对应的层与判定认定成立的层的数目差（第一栏）——无字段。跨次直接比较（第二栏）——无，病历按时间流水书写，没有任何一处要求与首诊并排。方外加减针对的特征（第三栏）——无，加减写在处方里，理由不写。复核所依据的证据项日期构成（第四栏）——无。未采纳意见（第五栏）——无，本院复核记录只有结论栏。无从判定（第六栏）——无，诊断栏不允许空着，填写率是被考核的。分类变更的当日观察项（第七栏）——无。观测项的共现结构（第八栏）——无。

一栏为“有而不在判定主线上”。第九栏：本院参加过一项临床研究，方案书里既有诊断标准条目表，也有疗效评价条目表，两份表都在，**可以并排放，数得出交集**。这一栏因此是“有”。但那两份表在研究方案里，不在这位患者的诊疗流程里；主治医生在做每月那一次判定时，不会去看它们。**按规则一，这一栏记为“有而不在判定主线上”，不计入“有”**。

走查的结论是一句让人不舒服的话：**九栏俱空，而这是一份优秀病历**。这不是这家机构的问题——本书预测绝大多数机构的结果与此相同，因为九栏在现行的任何一套病历规范里都没有位置。

六·本章最脆的一处：编码者一致性没有做

二十七问看上去是客观的——数数与查有无，似乎不留判断余地。但至少有三处需要判断：

其一，“有自由文本区但不与结论同屏”这一档的边界。同屏是一个界面事实，而不同的信息系统对“同屏”有不同的实现（同一页面折叠区算不算？点一下展开算不算？）。

其二，第一栏的“层数”须由编码者从病历重构。一位编码者可能把“脾肾阳虚兼水湿”读成两层，另一位读成三层。第一章给了重构规则，但规则本身要人来用。

其三，第七栏的“新的观察项”须判断某一项是不是新的。换了一个名字的旧指标算不算新？

三处都可能造成编码者之间的分歧。而检验方法极其便宜：**两组编码者，各自独立编码同一批一百份记录，算一致性系数。**这比本书任何一条正面预测都便宜，且不需要任何机构改变自己的流程——用现成的存档记录就能做。

本书没有做。这是本书欠的第二笔账，导论已认过一次，这里是它的具体位置。在做出来之前，二十七问只能用于探索，不能用于任何比较——不能比机构，不能比科室，更不能比人。一致性门槛沿用通行做法：低于 0.60 不可用；0.60 至 0.75 仅探索性使用；高于 0.75 方可用于群体层面的比较；**未经伦理审查，任何情形下不得用于对个体的处置或评价。**

七·三种误用，各配一条防线

误用一：做成检查表。最可能的一种。把九栏写进病历规范，把填写率纳入考核。**结局是九栏俱在而九栏俱空——每一栏都设立了，每一栏都填“无”“无异议”“不适用”。这不是危言耸听，它是规则三的直接后果。防线：验收时看的是九个读数真算出来的数，不看有没有设立这些字段。**一个机构可以做到九栏齐备而九个读数全部为零，那正是本书预测的失败方式。

误用二：用于评价人。把弃判率低读成“这位医生水平高，很少判不了”，或把它读成“这位医生不敢承认判不了”。两种读法都把一个制度读数按到了个人身上。**防线：九个读数的分母全部是判定次数或项数，不是人；报告时不得按医生分组。**第十四章的五条伦理约束里，这一条排在第一。

误用三：把“有而不在判定主线上”记成“有”。这是最容易发生也最难察觉的一种——它使一个机构可以靠“我们研究方案里有”“我们质控报表里有”把九栏全部报成有，而临床判定的那一次仍然什么也没变。**防线：规则一是硬的，不设例外。**若一个字段不在被判定者那一份记录上，它就不在。

八·这份问卷不承诺什么

它不测判定的准确性。它不测疗效。它不测医生的水平。它不告诉任何人应该怎么辨证。

它只回答一个问题：这一次判定做了什么，账上留下了什么。九栏俱空，并不意味着这次判定错了——第五节那份档案里的每一次判定都可能都是对的。它意味着的是另一件事：若这次判定后来被证明错了，没有任何东西可以被回查。

这句话是本书全部实践主张的最小形式，也是它唯一敢承诺的东西。

第十三章 九栏之间是相乘还是相加

一·一个不能回避的问题

九章各自指认了一笔被吸收的量。假定九栏都立起来了、九个读数都能算，还有一件事没有回答：它们合起来构成什么？

这不是一个纯理论的问题。它有一个非常具体的后果：若一个机构只能改一处，该改哪一处，改了会有多大用。若九栏是相加的，改任何一处都得到大致相当的一份改善，从最便宜的那处下手即可；若是相乘的，改一处几乎没有用——只要其余各处还在零附近，乘积仍然在零附近。

本章的主张是相乘，但主张的强度必须先说清楚：这是一项待检验的假说，不是一个已经被数据支持的模型。本章只有次序，没有数据。

二·三类关系，不是一条链

第一本书把九栏排成了一条链。本书不能照办，因为本书九栏之间的关系不是一种，是三种。

第一类：串联，相乘。留底率、存异率、弃判率三项。它们共同支撑一件事——这一次判定后来若被证明有问题，能不能回查。回查需要三样东西同时在场：一份未被这次处置改写过的证据（留底）、一份当初异议的技术内容（存异）、以及一个“当初本来就判不了”的记号（弃判）。任一为零，回查即断。三者相乘而非相加，第五章已经在与第四章的划界里给过一个具体的数：留底率 0.1 与存异率 0.1 的一次复核，其可回溯性不是 0.2，而是接近 0.01。

第二类：并联，取较大者。让层比与配格率。两者指向的是同一件事的两个记录位置——“这一次的处置只覆盖了对象的一部分”。让层比从判定那一侧记（认定成立而未动的层），配格率从处方那一侧记（为配上这一个人而做的调整）。记下任何一个，那件事就有了痕迹；两个都记，得到的不是双份，是同一件事的两个视角。因此这一对不相乘，取较大者。这也是第十一章预测它们相关系数在 0.3 至 0.6 的原因。

第三类：总开关，阈值。填判一项。它不是链上的一环，它在链的上游之外。理由是：其余八个读数都在问“这一次判定做了什么”，而填判问的是“这一次判定该不该发生”。若一次判定本该记为无从判定而被填成了判定，那么其

余八项算出来的一切都是在描述一次不该发生的判定的内部构造。让层比再漂亮，也只是漂亮地记录了一次不该做出的分层。

因此本章给出的合成形状不是一个连乘式，是这样一句话：

可回溯性 = [这一次判定是否本该发生] × [留底率 × 存异率 × 弃判率] × [max (让层比记录, 配格率记录)]，其余三项（吞差率、随程率、消参率）不在这条式子里——它们量的不是这一次判定的可回溯性，是跨次与跨轮的可比性，属于另一条轴。

九栏不在同一条式子里，这是本章相对第一本书最重要的一处修正。第一本把九栏排成一条链，本书发现自己的九栏分属两条轴：单次判定的可回溯性（六项）与跨次跨轮的可比性（三项）。两条轴各自成立，且可以一高一低——一个机构可以在单次上完全不可回溯而在跨次上高度可比（正是第二份档案里那份二十四次判定一致的优秀病历），反之亦然。

力一系轴：跨次跨轮的可比性。 台左半 随程半 消参半 （个住上只内）

图七 三类关系：总开关、串联相乘、并联取较大者

三·四个双向量必须先做方向转换

九个读数里有四个是双向量，不能直接进任何乘积式或加和式：

让层比高不必然坏。它高，说明这一次判定诚实地承认自己只动了一层。第一章明写过它自带的那件伦理：若把让层比压低设为目标，压低它最省事的办法是不再认定那些层成立——那正是这个读数所要防的事。

配格率同理，第三章写死了“不得设为考核指标，无论正向负向”。

弃判率是四个里最危险的。它为零说明体系不允许说“判不了”；它过高说明推诿。这个读数没有一个“越靠近某个值越好”的说法，只有一个区间，而这个区间本书给不出。

随程率高说明分类在跟着程序走，低说明分类稳定——但稳定也可能只是没人再修订标准了。

其余五个方向明确：留底率、存异率越高越好；吞差率、消参率、同据率越低越好。

因此任何合成式的第一步都不是加权，是为四个双向量各定一个参照区间，把读数转换成“距参照区间的偏离”。而这四个参照区间，本书一个也给不出——它们必须由实测分布决定，而实测一次也没做。这使本章的合成式在当前状态下不能被计算，只能被讨论。这句话应当被读成对本章的削弱，而不是一句谨慎的套话。

四·相乘能解释、相加解释不了的那件事

一个模型值不值得留着，看它能不能解释一件竞争模型解释不了的事。相乘能解释这一件：

为什么“加一个字段”在试点里有效，推广之后无效。

试点通常是这样做的：选一家配合度高的机构，派人驻点，把该补的记录一并补齐——留底补了，异议记了，判不了的也允许记了。**试点补的是多个环节。**结果好，于是写成经验推广。推广时能落地的通常只有一条——最容易写进规范的那一条，多半是“增设未采纳意见栏”。

相加模型预测：推广后应当得到试点效果的一部分，大致按补上的环节数打折。**相乘模型预测：几乎得不到任何效果。**若三项串联，把存异率从 0.1 提到 0.8 而留底率仍是 0.1、弃判率仍是 0，乘积仍是零——因为其中一项是零，乘积就是零，另外两项提得再高也没有用。

这个分歧是可判决的。**判决形式：在一批实施了单一字段改造的机构里，控制实施保真度之后，“结论被推翻后能否查出当初争议”的改善幅度，是否随其余两项读数的初始值变化。**相加模型预测不随，相乘模型预测强烈依赖。这一条不需要九项齐备，只需要三项，且三项都是可从既有归档件清点的 D 型或近 D 型读数。这是本书最便宜的一次判决性检验，成本远低于第十一章那张完整相关矩阵。

五·分层的做法，以及它可能否定整个第四编

上一节的判决检验有一个前提：三项串联这件事本身成立。若它不成立，本章连同第十一章都要重写。因此必须给出一个不预设答案的做法。

正确的做法是同时拟合四个模型，用样本外预测或信息准则比较，而不是先信一个再去找证据：

可加模型：可回溯性 = 各项的加权和。**交互模型：**加和再加两两交互项。**乘积模型：**各项之积。**瓶颈模型：**取各项的最小值。

四个模型对同一批数据会给出不同的样本外预测，且分得开：可加模型预测单点改造有部分效果；乘积与瓶颈都预测几乎无效，但两者在“两项都不为零时”分岔——乘积对第二项的提升敏感，瓶颈不敏感。**因此三个模型之间，只需要一批同时测了三项、且三项取值有足够散布的机构数据就能分开。**

必须写下这句话：**第一种结果会否定整个第四编。**若可加模型胜出，那么九栏之间没有结构关系，它们只是九个独立的记录缺口；第十一章的相关预注册、本章的三类关系、第十二章把九句合成一份问卷的做法，全都失去理由——九章仍然各自成立，但它们不构成一本书，只构成一个文集。**本书接受这个结果，并在此写明：若那样，本书应当被拆回九篇独立论文，第四编整个作废。**

六·乘积模型给了本书一条退路，本章自己堵上

相乘模型有一个使用者不会主动提起的好处：**它对任何失败都自带解释。**某机构补了字段而没有改善？因为别处还是零。补了两处仍没改善？因为第三处还是零。补了三处仍没改善？因为方向转换没做对。这条退路可以一直退下去，退到本章永远不会错。

一个永远不会错的模型不是一个模型。因此在这里预先堵死：

在一家基线三项均低的机构，若把留底率从 0.1 提到 0.8 而其余两项保持不动，“结论被推翻后能否查出当初争议”的改善幅度达到或超过三倍，则乘积模型在该场所被否，本章须改用可加或交互模型重写。

三倍这个数不是随便取的：乘积模型预测的改善幅度约等于该项的提升倍数乘以其余项的乘积，在其余两项接近零时应当接近于无；可加模型预测该项提升八倍应带来与其权重成比例的一份可观改善。**三倍是两者预测之间的一条可以落笔的线。**它可能取得不准，但它使本章可以在一家机构、一次改造、一批既有归档件上被推翻。

七·本章的位置

本章不新增读数，不新增判据。它给出的是一个合成形状与一条可以推翻它的观察。

按外部标准，本章的成色是三章里最弱的：三类关系的划分基于各章的论证而非数据，四个参照区间给不出，合成式在当前状态下不能被计算。**本章的正确读法是一份待检验的设计说明，不是一个结论。**引用本章时若把它当作“本书主张九栏相乘”这样一个结论来引，是对本书的误用——第十四章会说明，这类误用是本书预测会大量发生的三种之一。

第十四章 结算的反噬与伦理总则

一·这一章为什么必须写在书里而不是附录

九章指认了九笔被吸收的量，第四编前三章说了怎么算、怎么问、怎么合成。剩下最后一件事：**把这九栏立起来之后，会引出什么新麻烦。**

这件事不能放在附录，理由与第一本书相同而更迫切：**本书的处方比上一本更容易反噬。**上一本要求的是给一栏不存在的东西设一个位置——设了就设了，那样东西本来就没有被谁在做。本书要求的是把已经发生、已被吸收的那一份从结论里剥出来，而**那一份的产生者往往是一个在场的人**：那位提了异议的进修医生，那位判不了却必须填的医生，那位在方里加了两味药却不写理由的医生。**剥出来意味着把人的判断行为变成可归档的对象**，这件事一旦做起来，会改变判断行为本身。

二·三种反噬

第一种：字段有而被填成“无异议”。这是最普通、也最可能的一种。九栏写进规范，填写率纳入考核，于是每一栏都设立了，每一栏都填上一个使自己免于追问的值——未采纳意见栏填“无异议”，无从判定栏一年零次，方外加减的理由栏填“随证加减”。

这一种在第十二章已经预告过，它是编码规则三（只记有无、不记内容质量）的直接后果。**本书接受这个后果而不放宽规则三**，因为一旦开始评判内容质量，问卷就重新需要评判者，而九句判据的全部价值在于不需要评判者。

第二种：留痕的动作改变了被留痕的东西。这一种是本书特有的，也最难解。

一旦所有在场者都知道“未采纳意见会被归档”，会发生两件相反的事，而且很可能同时发生。**其一，异议减少。**一位进修医生本来会随口说一句“我看着像湿热”，现在他知道这句话会进档案、会被日后调阅、可能会被读成他与上级不一致的记录，于是不说了。**分歧被记录的制度，减少了分歧的产生。**其二，异议变得程式化。仍然会有人提，但提的内容会向“安全的异议”漂移——提那些日后被证明错了也无所谓的，不提那些真正要紧而自己没把握的。

这不是激励设计不当，改一改就能修好。它的根子在于：**这一笔量之所以携带信息，恰恰因为它产生于一个不留痕的场合。**那十分钟的分歧之所以有临床价值，一半来自它是自由说出来的。把它变成归档对象，就把它从“自由说出来的判断”变成了“要负责的书面意见”，而这是两种东西。

第三种：弃判被优化。这是本书唯一一个“设立一个新输出格式”的处方，也是九个里最容易被滥用的。第六章自己已经说过这一点。

一旦弃判率成为一个被鼓励的指标——比如被读作“这家机构敢于承认自己的边界”——立刻会出现战略性弃判：难病例记为无从判定，责任随之转出。而这恰恰与第六章的处方方向相反。第六章要的是让体系**有能力**说“这里不该由我来判”，不是让它**有动力**这么说。

图八 三种反噬及其对策

三·三条不彻底的对策

对策一，针对第一种：验收看真算出来的数，不看有没有设立字段。一个机构可以做到九栏齐备而九个读数全部为零；验收若看字段设立，这个状态就是满分。第十二章已立此条，此处只补一句执行细节：“**未采纳意见栏非空份数占比**”与“**未采纳意见栏存在与否**”必须分开报告，且前者不得被后者替代。

对策二，针对第二种：留痕的时点必须在结论之前，不在结论之后。这是本章唯一一条有实质结构的对策。

事后追记“未采纳意见”，记录的是一个人在知道自己输了之后仍然坚持的东西——那是一份态度声明。**在结论形成之前先登记各在场者各自的判断，记录的是过程状态。**后者不要求任何人“提出异议”，只要求每个人写下自己看到了什么；分歧是事后从这些记录里读出来的，不是由谁主动“提”出来的。

这一条能减少但不能消除第二种反噬——人仍然知道自己写下的东西会被存下来。本书**不主张它能解决问题，只主张它比事后追记损害小。**这个判断本身是可以被检验的：在两种记录时点下各跑一年，比较所记分歧的数量与内容分布。本书没有做。

对策三，针对第三种：弃判必须带下游处置路径。第六章的弃判率定义里本来就有这个限定——记为“不适用／无从判定”且带有下游处置路径的次数。本章把这个限定词从一个技术细节升为一条硬防线：

一次弃判若不指向任何下一步（转诊、会诊、延判并约定复查时点、或转由另一套判据处理），它不计入弃判率，且不构成一次合格的弃判。

有了这一条，战略性弃判的成本就不再为零——推诿者必须同时指出谁来接。这不能杜绝推诿，但它把推诿从一个免费动作变成一个要签字的动作。

四·五条伦理硬约束

以下五条不是建议，是使用本书九栏的前提条件。它们已被提到使用声明页，此处给出完整表述与理由。

第一条：九个读数指的是判定的位置，不指人。报告不得按医生分组、按个人排名，不得进入任何形式的个人绩效。理由：九个读数的分母全部是判定次数或项数，不是人；把一个以判定为单位的读数按人聚合，得到的数没有可解释的含义，而它会被当成有。

第二条：不得追溯问责历史判定，填判尤其不得。一次填判，在一个没有“无从判定”这个输出格式的体系里，是当时唯一可能的输出。责任在体系不在填的人。任何把本书用于回溯追责的做法，都会立刻使第一种反噬变成必然——所有人都会开始把历史记录填成安全的样子。

第三条：编码者一致性做出来之前，不得据本书作任何不利于机构或个人的安排。门槛沿用第十二章：低于 0.60 不可用，0.60 至 0.75 仅探索性使用，高于 0.75 方可用于群体层面的比较。本书自己没有做这一步，因此按本条，本书当前不支持任何不利安排。

第四条：弃判权不得与绩效挂钩，无论正向还是负向。正向奖励导致推诿，负向惩罚导致填判——两个方向各自导向第六章所指认的两种病。这一条与第三章“配格率不得设为考核指标，无论正向负向”是同一个形式，写在两处。

第五条：未采纳意见的记录，不得对被记录者以外的第三方开放用于评价。它记录的是一位在场者的专业判断，不是他的错误记录。可用于回查、可用于研究（去标识后）、可用于教学（经本人同意），不可用于评价这个人。这一条若

被违反，第二种反噬会在一到两年内把这一栏彻底掏空，而且不可逆——一旦一个群体学会了不在那里写真话，重新学会需要的时间远长于学会的时间。

五·本书预言自己的结局

按上述三种反噬的相对强度，本书对自己最可能的结局作出如下预言，写死在这里以便日后核对：

九栏中会有一到两栏被写进某一版病历书写规范或临床研究报告规范——最可能的是未采纳意见栏与无从判定栏，因为它们最容易被表述成一条填写要求。其余七栏不会。理由：其余七栏都要求改变的是记录的结构（把处置那一侧与对象那一侧分开、把条目表并排列出、把归属变更与观察项分开记），而不是增加一个填写项，改动成本高一个量级。

被写进去的那一两栏，会在三到五年内被填成空的。填写率会很高，非空率会很高，而技术内容会趋近于零——“无异议”“无”“不适用”。

届时会出现一批研究报告这一两栏填写率的文献，用它来说明中医病历质量的改善。这将是本书被误用得最彻底的一次，而本书对此没有任何阻止手段，只有一条事先写下的判据：

验收看的是真算出来的读数，不是字段的设立率，也不是字段的填写率。

这句话在本书里出现了三次（第十二章、本章第三节、此处）。三次不是重复，是因为它是本书唯一一条能在误用发生之后仍然被引用的防线。

六·本书应当被整体撤销的条件

一本书应当说清自己在什么情况下应当被整体撤回，而不只是某一条推论被修正。本书的条件有三，任何一条成立即足够：

其一。若第十三章第五节的四模型比较中，可加模型胜出——九栏之间没有结构关系，只是九个独立的记录缺口——则第四编作废，本书应被拆回九篇独立论文，“结算学”这个名字应当被撤销，因为它主张的正是这九笔量属于同一次结算。

其二。若在一批已经把九栏全部立起来、九个读数全部落在合理区间的机构里，判定质量在两年内没有可测的改善，则本书的实践主张失败。**这一条不撤销**

诊断主张——九笔量仍然可以是真实存在且未被记录的，只是记录它们不带来改善。但一本以“结算”为名的书，若结算完了什么也没变，它应当把书名改掉。

其三。若第二种反噬被证明是压倒性的——即在所有已知的记录时点设计下，留痕都使被留痕的那类判断显著减少且内容质量下降——则本书的处方在净效果上有害，应当整体撤回。**这一条是本书自己最担心的一条**，因为它不依赖本书的任何论证出错：九笔量可以全部真实存在，九个读数可以全部算得出来，而把它们记下来这件事本身仍然是净损失。

第三条与前两条不同：前两条错在论证，第三条错在处方，而处方错的时候论证可以是对的。**本书没有能力事先排除第三条。**它只能把它写在这里，并把对策二（记录时点前置）作为一条可以被检验的减损方案交出去。

七·一句总则

九章、九个读数、二十七问，全部服从同一条总则：

结算的目的不是让被吸收的一切都吐出来，而是让每一笔被吸收的量，都有人知道它被吸收了。

知道它被吸收了，与把它记下来，是两件事。前者是本书的最低要求，后者是本书的处方。**在处方被证明有害的场合，最低要求仍然成立。**这也是本书与一切“提高记录完整性”的主张之间最后一条分界。

结语 三份档案，第二遍

一·回到三份档案

导论开篇摆了三份档案。现在把它们重看一遍。

第一份，那位七十岁老人，三位医生辨出三个证，三张方各自有效。第一遍读它的时候，可用的解释只有两种：辨证水平的差异，或者殊途同归。九章之后，有了第三种读法。三位医生看见的层是同一批——畏寒、肢冷、腰膝酸软、夜尿频、舌淡胖、脉沉细，这些层在三个人那里都被认定成立。**他们的分歧不在“认定哪几层成立”，在“这一次动哪一层”**。三张方各自有效，因为每一张都动对了一层；三个证名不同，因为证名记的是被动的那一层，而不是被认定成立的全部。**病历上那四个字从来没有承诺过它记的是全部——是我们以为它记的是全部**。这一份档案里被吸收的量，是让层。

第二份，那份二十四次判定一致的优秀病历。第一遍读它，它是一份可以拿去评优的记录。九章之后，它是这本书里最难堪的一份材料：**正因为它每一次都判得对，它才把一整段真实的变化吞掉了**。二十三次“与上次比没有明显变化”，每一次都是一位有经验的医生在两次之间做出的正确判断；而首末两端确实不是同一个证。没有任何一次可以被追责，因为没有任何一次是错的。这一份档案里被吸收的量，是吞差，而它之所以能被吞得这么干净，正是因为这家机构的记录制度足够规范。

第三份，那十分钟。第一遍读它，它是一次正常的会诊。九章之后，它是九笔量里唯一一笔有名有姓的持有者的量——那位进修医生在场，他能说出依据，他的判断内容不在任何一份可检索的文件里。导论第七节说过，这正是本书与上一本书的分界所在：**缺栏没有持有者，吞并有**。而第十四章说过，正因为它有持有者，把它记下来这件事会反过来改变他下一次说不说话。这一份档案里被吸收的量，是摊差，而它是九笔里最难被安全地结出来的一笔。

三份档案，没有一份是失误。这句话在导论里写过一次，现在它有了具体的内容：不是“没有人犯错”，而是**这三份档案里的每一个动作，都是在套没有这九栏的账上，一个人能做出的最好的动作**。

二·结算学真正的任务

写到这里，必须把本书的任务重新表述一次，因为它与开篇的表述已经不同。

开篇说：结算是把已经发生、被结论吃掉的那一份从结论里剥出来。这句话没错，但它省略了一件要紧的事——**不是每一笔被吸收的量都应当被剥出来。**

九章之后，加上第十四章那三种反噬，本书真正的任务是区分三种吸收：

第一种，应当被结出来的。它有一个曾经在场的持有者，有一个明确的下游用途，剥出来的成本低，而剥出来的动作不改变产生它的那个动作。**留底率、同据率、弃判率所对应的三笔属于这一类**——记下证据项的日期、把两份条目表并排列出、给“判不了”设一个格，这三件事都不改变判定行为本身。

第二种，只能被群体审计的。个案层面不可回溯，但分布层面可查。**吞差率、随程率、消参率属于这一类**——你无法对某一位病人说“你这二十四次里被吞掉了多少”，但你可以在一批病历上算出这个体系的吞差率；你无法指出某一次分类变更是不是被程序推动的，但你可以数最近 N 次里有几次当日没有新的观察项。这一类的正确用法是诊断制度，从一开始就不该落到个案上。

第三种，必须被允许继续吸收的。若把它结出来，就毁掉了产生它的那个动作。**摊差里的一部分属于这一类。**一次会诊中的自由讨论，其临床价值有一半来自它是自由说出来的；把每一句异议都变成归档对象，得到的是一份完整的记录和一场再也不会发生的讨论。

三类的判据只有一条，而且它与第一本书那条是同一形式的下一步：

当把一笔量结出来的动作，会改变或损害产生这笔量的那个动作时，它属于第三类。

这条判据难用。它要求判断“这个动作会不会被改变”，而这个判断在做出来之前无法验证。本书只能补一句执行上的约束：**这个判断只能由承担那个动作的一方作出，且必须可撤回。**一位主治医生可以说“这一类讨论我们不记”，一位院长不可以替他说；说过之后，可以改回来。

这把本书从一本“要求增加记录”的书，改写成一本关于**哪些量该结、哪些量只能群体审计、哪些量必须被允许继续吸收的书**。这是本书愿意被记住的形式。

判据：结出来的动作若改变或损害产生它的动作，属第三类

图九 三种吸收

三·三个洞

第一个洞：一次都没跑。 九句判据里最便宜的三条——把诊断标准的条目表与疗效评价的条目表并排放数交集、数一段时期的判定输出里有几次记为无从判定、看一份复核所依据的证据项各自的日期——一个下午可以做完。**本书一次也没有做。**这句话在导论里写过，在第十一章第八节写过，在这里写第三次，因为它是本书与“一本可以被相信的书”之间距离最短的一处。

第二个洞：编码者一致性没做。 两组人各自独立编码一百份既有归档记录，算一致性系数，不需要任何机构改变流程。本书没有做。在做出来之前，二十七问只能用于探索，**本书自己给出的伦理第三条因此适用于本书自己：**本书当前不支持任何不利于机构或个人的安排。

第三个洞：说不出该由谁牵头。 本书的处方要求改的是两样东西——病历书写规范与临床研究报告规范。而这两样的修订权，不在任何一位读过本书的临床医生手里，也不在作者手里。**本书能说出该改什么，说不出该由谁改。**这个洞的形状很眼熟：它正是本书第六章所指认的那件事——一个本该有归属的位置上，没有人被指派。本书的处方落在自己指认的缺栏里。

四·最后一句

九章合起来只主张一件事，而这件事可以不用本书的任何一个新词说出来：

每一次判定都做了一些事。它认定了哪几层成立，指定了这一次动哪一层，把这个人放进了一个格，在格里为他调了几味药，与上一次比了一次，吞下了一份差异，在复核时消灭了一份分歧，在不该判的时候仍然给出了一个判定，在跨轮次时重划了自己的对象，在建模时并掉了几个参数，并且用同一批条目既做了判定又做了验证。

这些事全都发生了。账上留下的，是一个证名。

本书要求的不是别的，就是让这句话里的每一个“做了”，在某处留下一个可以被数出来的痕迹——或者，在留痕会毁掉那个动作的地方，至少让人知道它被吸收了。

参考文献与文献线索

各章的文献按章排列，不合并去重——同一处文献在不同章里承担的功能不同，合并会使读者无法看出某一条究竟支撑了哪一章的哪一步。

凡标注为线索者，为按出现次序列出的可核查方向，非规范著录；引用者请核对原始文献，若有出入以原文为准。

第一章 让层

本章引用的判断分属不同传统，为便于核查，按出现次序列出主要线索。

关于作品同一性与记谱法判据：Nelson Goodman, **Languages of Art: An Approach to a Theory of Symbols**, Bobbs-Merrill, 1968, 尤其第三、四、五章关于 autographic 与 allographic 的区分，以及关于摹本与原作在感知不可区分时仍然不同的论证。

关于化学选择性与保护基：任何一本高等有机合成教材的保护基章节；关于无保护基路线的争论，参见 R. W. Hoffmann, “Protecting-Group-Free Synthesis”, **Synthesis**, 2006; P. S. Baran et al., “Total Synthesis of Marine Natural Products without Using Protecting Groups”, **Nature**, 2007; 关于步数经济性，参见 P. A. Wender 关于 step economy 的系列论述。关于产物分布与反应物构象分布的脱钩，参见任何物理有机化学教材关于 Curtin-Hammett 原理与 Winstein-Holness 方程的处理。

关于认知诊断评估：K. K. Tatsuoka 关于规则空间模型的工作（1983 起）；J. de la Torre 关于 DINA 模型参数估计的工作（2009）；以及关于 Q 矩阵设定与属性划分依据的方法学讨论。关于情境认知：J. S. Brown, A. Collins & P. Duguid, “Situated Cognition and the Culture of Learning”, **Educational Researcher**, 1989; J. Lave & E. Wenger, **Situated Learning**, 1991。关于效度与后果：S. Messick 关于 validity 的经典论述（1989）。

关于文物修复的最小干预、可逆性与可识别性：C. Brandi, **Teoria del restauro**, 1963; 《威尼斯宪章》1964。关于现状病害报告与修复处理报告的分立，参见各国文物保护档案规范。

关于技术债：W. Cunningham, “The WyCash Portfolio Management System”, OOPSLA 1992 experience report。关于延期维护欠账的会计处理, 参见各国公共设施管理与审计部门关于 deferred maintenance backlog 的报表规范。

关于正式结构与实际活动的脱耦：J. W. Meyer & B. Rowan, “Institutionalized Organizations: Formal Structure as Myth and Ceremony”, *American Journal of Sociology*, 1977。

关于中医方面：《素问·标本病传论》关于标本缓急的论述；《难经》关于传变的论述；现行《中医病历书写基本规范》与各类证候诊断标准；证候积分法在中药新药临床研究指导原则中的规定；以及中医体质学关于体病相关与体质可调的主张。

字数：正文约二万二千汉字。

版本：第一稿。第五节之一按第十三节第三处削弱重写过一次；核心主张的适用范围在成稿过程中由“任何判定”收窄为“处置手段选择性低于对象分节精度的系统”。

人机分工声明：本章的问题、三个学科的指定、以及成文要求由作者提出；材料梳理、论证组织与文字撰写由作者与人工智能协作完成；全部判断、削弱的接受、赌注条款与不算命中的五条，由作者定稿并负责。文中所有可核查的文献线索需读者自行复核，若有出入以原始文献为准。

第二章 吞差

本章引用的判断分属不同传统，为便于核查，按出现次序列出主要线索。

关于潜变量的实在性与局部独立性判据：心理测量学中关于反射性模型与形成性模型的争论；D. Borsboom 等关于 latent variable realism 与构念效度的论述（2003 年前后的系列工作）；潜类分析中关于互斥穷尽分割的标准假定。

关于粗粒化、有效理论与宏观态作为等价类：统计力学教科书中关于宏观态与微观态关系的标准处理；玻尔兹曼熵的组合解释；关于 Gibbs 佯谬与“可分辨性依赖于观察手段”的长期讨论。

关于可诊断性与不可区分故障：M. Sampath 等关于离散事件系统可诊断性的工作（1995）；卡尔曼可观测性判据；FMEA 与状态监测中关于可区分性的工程实践。

关于不可分辨关系的不传递性：心理物理学关于最小可觉差的经典传统（Weber、Fechner 一系）；R. D. Luce 关于半序的工作（1956）；N. Goodman 在《表象的结构》（1951）中关于 matching 关系不传递的论述；关于 tolerance space 的数学处理。

关于概念漂移与固定基准集：机器学习中 concept drift 检测的标准文献。关于电子病历的复制前带与记录膨胀：医学信息学中关于 copy-forward 与 note bloat 的实证研究。关于模糊集：L. A. Zadeh (1965)。关于测量不变性：心理测量学中关于纵向不变性检验的标准程序。

关于未更正错报的汇总：各国审计准则中关于重要性水平与未更正错报汇总表的规定。关于生长曲线与百分位追踪：儿科生长监测的标准做法与世界卫生组织生长标准。关于临床总体印象量表中“较基线变化”的评定要求：CGI 量表的标准说明。

关于语言学的链移：元音大推移与拉波夫一系关于进行中的音变的研究。

关于中医方面：《伤寒论》六经传变；卫气营血辨证；现行中医病历书写规范与各类证候诊断标准；证候疗效评价中的证候积分法；辨证一致性研究中关于 Kappa 系数的报告。

字数：正文约二万二千汉字。

版本：第一稿。第五节之一在成稿过程中增补了“吞差不是模糊”一条；第八节的比率定义由最初的“被吞差异之和占总差异之比”改为“(k-m)/k”，因为前者不可清点；主张的适用范围由“任何判定系统”收窄为“有时间序的逐次比较判定系统”。

人机分工声明：本章的问题、三个学科的指定与成文要求由作者提出；材料梳理、论证组织与文字撰写由作者与人工智能协作完成；全部判断、三处削弱的接受、赌注条款与不算命中的五条，由作者定稿并负责。文中所有可核查的文献线索需读者自行复核，若有出入以原始文献为准。

第三章 配格

本章引用的判断分属不同传统，为便于核查，按出现次序列出主要线索。

关于细胞类型的同一性判据与调控核心：单细胞测序时代关于 cell type 与 cell state 之争的综述；D. Arendt 等关于细胞类型由核心调控复合体定义、由进化同源性裁定的论述（2016 年前后）；关于成纤维细胞位置记忆与 HOX 基因位置编码的实验工作。

关于互换性制造与公差：机械设计与制造工艺教材中关于公差配合、互换性原则的标准处理；关于分组选配、配对件与“成对更换不得单换”的装配工艺规范；关于统计公差分析的工程方法文献。

关于计量口径与可比性：管理会计中关于一致性与可比性的原则；组织研究中关于绩效指标跨单位可比性的讨论；G. Bowker & S. L. Star, *Sorting Things Out: Classification and Its Consequences* (1999) 关于分类与标准的社会学考察；S. L. Star & J. Griesemer 关于边界物的工作（1989）。

关于可替代性：经济学中关于 fungibility 的标准区分。关于可读性与地方性知识：J. C. Scott, *Seeing Like a State* (1998)。关于隐性知识：M. Polanyi, *The Tacit Dimension* (1966)。关于自然类：科学哲学中关于 natural kinds 的标准讨论。

关于交叉配血：输血医学中关于血型鉴定与交叉配血作为两个独立必做步骤的规程，以及同型输血溶血反应促成交叉配血制度化的历史。关于器官移植配型：HLA 配型与配型分数系统的标准说明。关于治疗药物监测与个体化给药：临床药理学中关于治疗窗与剂量调整的标准做法。

关于大规模定制：B. J. Pine, *Mass Customization* (1993)。关于精准医学与分层医学的方法学讨论：肿瘤靶向治疗中关于分子分型与组内异质性的综述。

关于中医方面：《素问·异法方宜论》三因制宜；方证相应一系的著作；证素辨证的方法学文献；现行中医病历书写规范与证候诊断标准；中药新药临床研究指导原则中关于证候疗效评价与固定处方设计的规定。

字数：正文约二万二千汉字。

版本：第一稿。第八节之二的第二条限制（不得设为考核指标，无论正向负向）是在第十三节第一处削弱确立之后才写下的，此前本章原打算主张配格率越高越好；主张的适用范围由“任何分类系统”收窄为“处置手段对类内残余差异敏感且个体差异连续多维的系统”。

人机分工声明：本章的问题、三个学科的指定与成文要求由作者提出；材料梳理、论证组织与文字撰写由作者与人工智能协作完成；全部判断、三处削弱的接受、赌注条款与不算命中的五条，由作者定稿并负责。文中所有可核查的文献线索需读者自行复核，若有出入以原始文献为准。

第四章 并肩

- 国际标准化组织：测量方法与结果的准确度（正确度与精密度）相关标准；实验室间比对能力验证的统计方法相关标准。
- 人用药品注册技术要求国际协调会：分析方法验证相关技术指导原则。
- Gy, P. 取样理论相关著述。
- Messick, S. (1989). Validity. In *Educational Measurement* (3rd ed.).
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning. *Psychological Science*.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful Retrieval Attempts Enhance Subsequent Learning. *JEP: LMC*.
- Brandi, C. (1963). *Teoria del restauro*.
- Goodman, N. (1968). *Languages of Art*.
- Rembrandt Research Project, *A Corpus of Rembrandt Paintings*, vols. I–VI.
- Senn, S. (2002). *Cross-over Trials in Clinical Research* (2nd ed.).
- Grizzle, J. E. (1965). The Two-Period Change-Over Design and Its Use in Clinical Trials. *Biometrics*.
- Feinstein, A. R., Sosin, D. M., & Wells, C. K. (1985). The Will Rogers Phenomenon. *NEJM*.
- Illich, I. (1975). *Medical Nemesis*.
- Hacking, I. (1995). The Looping Effects of Human Kinds. In *Causal Cognition*.

- Merton, R. K. (1948). The Self-Fulfilling Prophecy. *The Antioch Review*.
- MacKenzie, D. (2006). *An Engine, Not a Camera*.
- Callon, M. (ed.) (1998). *The Laws of the Markets*.
- Power, M. (1997). *The Audit Society: Rituals of Verification*.
- Levinthal, D. A., & March, J. G. (1993). The Myopia of Learning. *Strategic Management Journal*.
- Leonard-Barton, D. (1992). Core Capabilities and Core Rigidities. *Strategic Management Journal*.
- 张仲景：《伤寒论》辨太阳病脉证并治相关条文。

第五章 摊差

- 国际计量局等：国际单位制相关文件；测量不确定度表示指南相关文件。
- CODATA Task Group on Fundamental Constants: 基本物理常数推荐值及其平差方法说明（历次版本）。
- Birge, R. T. (1932). The Calculation of Errors by the Method of Least Squares. *Physical Review*.
- Riess, A. G. et al.; Planck Collaboration: 哈勃常数测定与张力相关系列工作。
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*.
- Dalkey, N., & Helmer, O. (1963). An Experimental Application of the Delphi Method. *Management Science*.
- Janis, I. L. (1972). *Victims of Groupthink*.
- Rosenthal, R. (1979). The File Drawer Problem and Tolerance for Null Results. *Psychological Bulletin*.
- Polanyi, M. (1966). *The Tacit Dimension*.

- Collins, H. M. (1985). *Changing Order: Replication and Induction in Scientific Practice*.
- Kelly, T. P. (1998). *Arguing Safety: A Systematic Approach to Managing Safety Cases*.
- Leveson, N. G. (2011). *Engineering a Safer World*.
- Huber, P. J. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*.
- Levinthal, D. A., & March, J. G. (1993). The Myopia of Learning. *Strategic Management Journal*.
- Leonard-Barton, D. (1992). Core Capabilities and Core Rigidities. *Strategic Management Journal*.
- Merton, R. K. (1940). Bureaucratic Structure and Personality. *Social Forces*.

第六章 填判

- Chow, C. K. (1970). On Optimum Recognition Error and Reject Tradeoff. *IEEE Transactions on Information Theory*.
- Scheirer, W. J. et al. (2013). Toward Open Set Recognition. *IEEE TPAMI*.
- Quiñonero-Candela, J. et al. (eds.) (2009). *Dataset Shift in Machine Learning*.
- Regev, A. et al. (2017). The Human Cell Atlas. *eLife*.
- Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden. *Scientific Reports*.
- Luecken, M. D., & Theis, F. J. (2019). Current Best Practices in Single-Cell RNA-Seq Analysis. *Molecular Systems Biology*.
- Trapnell, C. (2015). Defining Cell Types and States with Single-Cell Genomics. *Genome Research*.
- Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*.

- Simon, H. A. (1997). **Administrative Behavior** (4th ed.).
- Campbell, D. T., & Stanley, J. C. (1963). **Experimental and Quasi-Experimental Designs for Research**.
- Cronbach, L. J. (1982). **Designing Evaluations of Educational and Social Programs**.
- Merton, R. K. (1940). Bureaucratic Structure and Personality. **Social Forces**.
- Levinthal, D. A., & March, J. G. (1993). The Myopia of Learning. **Strategic Management Journal**.
- Leonard-Barton, D. (1992). Core Capabilities and Core Rigidities. **Strategic Management Journal**.
- Welch, H. G., Schwartz, L., & Woloshin, S. (2011). **Overdiagnosed**.
- 国际电工委员会、国际标准化组织：型式试验与合格评定相关标准；环境试验条件与试验方法相关标准。
- 世界卫生组织：国际疾病分类相关文件中关于“未特指”类目的使用说明。

第七章 随程

Bridgman, P. W. (1927). **The Logic of Modern Physics**. New York: Macmillan.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. **Psychological Bulletin**, 52(4), 281–302.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). **The Dependability of Behavioral Measurements**. New York: Wiley.

Dickie, G. (1974). **Art and the Aesthetic: An Institutional Analysis**. Ithaca: Cornell University Press.

Farquhar, J. (1994). **Knowing Practice: The Clinical Encounter of Chinese Medicine**. Boulder: Westview Press.

Goodman, N. (1968). **Languages of Art: An Approach to a Theory of Symbols**. Indianapolis: Bobbs-Merrill.

Hacking, I. (1995). The looping effects of human kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), **Causal Cognition: A Multidisciplinary Debate**. Oxford: Clarendon Press.

Hacking, I. (2007). Kinds of people: Moving targets. **Proceedings of the British Academy**, 151, 285–318.

Leonard-Barton, D. (1992). Core capabilities and core rigidities. **Strategic Management Journal**, 13(S1), 111–125.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. **Strategic Management Journal**, 14(S2), 95–112.

Merton, R. K. (1940). Bureaucratic structure and personality. **Social Forces**, 18(4), 560–568.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), **Educational Measurement** (3rd ed.). New York: Macmillan.

Miller, D. (1990). **The Icarus Paradox**. New York: HarperBusiness.

Polanyi, M. (1958). **Personal Knowledge: Towards a Post-Critical Philosophy**. London: Routledge.

Polanyi, M. (1966). **The Tacit Dimension**. Garden City: Doubleday.

Rauret, G., et al. (1999). Improvement of the BCR three step sequential extraction procedure. **Journal of Environmental Monitoring**, 1(1), 57–61.

Rheinberger, H.-J. (1997). **Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube**. Stanford: Stanford University Press.

Scheid, V. (2002). **Chinese Medicine in Contemporary China: Plurality and Synthesis**. Durham: Duke University Press.

Selznick, P. (1949). **TVA and the Grass Roots**. Berkeley: University of California Press.

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. **Journal of Educational Measurement**, 30(3), 215–232.

Templeton, D. M., et al. (2000). Guidelines for terms related to chemical speciation and fractionation of elements. **Pure and Applied Chemistry**, 72(8), 1453–1470.

Tessier, A., Campbell, P. G. C., & Bisson, M. (1979). Sequential extraction procedure for the speciation of particulate trace metals. **Analytical Chemistry**, 51(7), 844–851.

沈自尹等：肾阳虚证的临床与实验研究系列报告，上海，一九五九年起。

陈可冀等：血瘀证与活血化瘀研究系列报告。

《中医虚证辨证参考标准》，全国中西医结合虚证与老年病研究专业委员会，一九八六年。

《中药新药临床研究指导原则》，一九九三年版及其后续修订。

《中医临床诊疗术语·证候部分》（GB/T 16751.2），一九九七年版及其后续修订。

朱文锋：《证素辨证学》，二〇〇八年。

第八章 消参

Åström, K. J., & Wittenmark, B. (1995). **Adaptive Control** (2nd ed.). Reading: Addison-Wesley.

Bellman, R., & Åström, K. J. (1970). On structural identifiability. **Mathematical Biosciences**, 7(3–4), 329–339.

Epstein, S. (1979). The stability of behavior: On predicting most of the people much of the time. **Journal of Personality and Social Psychology**, 37(7), 1097–1126.

Fisher, M. E. (1974). The renormalization group in the theory of critical behavior. **Reviews of Modern Physics**, 46(4), 597–616.

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). **Synthese**, 28(2), 97–115.

Kadanoff, L. P. (1966). Scaling laws for Ising models near T_c. **Physics Physique Fizika**, 2(6), 263–272.

- Kuhn, T. S. (1962). **The Structure of Scientific Revolutions**. Chicago: University of Chicago Press.
- Ljung, L. (1999). **System Identification: Theory for the User** (2nd ed.). Upper Saddle River: Prentice Hall.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. **Psychometrika**, 58(4), 525–543.
- Mischel, W. (1968). **Personality and Assessment**. New York: Wiley.
- Pearl, J. (1995). Causal diagrams for empirical research. **Biometrika**, 82(4), 669–688.
- Polanyi, M. (1958). **Personal Knowledge**. London: Routledge.
- Polanyi, M. (1966). **The Tacit Dimension**. Garden City: Doubleday.
- Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), **Art, Mind, and Religion**. Pittsburgh: University of Pittsburgh Press.
- Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. **American Journal of Orthopsychiatry**, 6(3), 412–415.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, 1(5), 206–215.
- Villaverde, A. F., Barreiro, A., & Papachristodoulou, A. (2016). Structural identifiability of dynamic systems biology models. **PLOS Computational Biology**, 12(10), e1005153.
- Walter, E., & Pronzato, L. (1997). **Identification of Parametric Models from Experimental Data**. Berlin: Springer.
- Wampold, B. E. (2001). **The Great Psychotherapy Debate**. Mahwah: Lawrence Erlbaum.
- Wilson, K. G. (1971). Renormalization group and critical phenomena. **Physical Review B**, 4(9), 3174–3183.
- 沈自尹等：肾阳虚证的临床与实验研究系列报告，上海，一九五九年起。

陈可冀等：血瘀证与活血化瘀研究系列报告。

《中医虚证辨证参考标准》，全国中西医结合虚证与老年病研究专业委员会，一九八六年。

《中医临床诊疗术语·证候部分》（GB/T 16751.2），一九九七年版及其后续修订。

朱文锋：《证素辨证学》，二〇〇八年。

第九章 同据

Brogden, H. E., & Taylor, E. K. (1950). The theory and classification of criterion bias. **Educational and Psychological Measurement**, 10(2), 159–186.

Burnet, F. M. (1957). A modification of Jerne's theory of antibody production using the concept of clonal selection. **Australian Journal of Science**, 20, 67–69.

Campbell, D. T. (1979). Assessing the impact of planned social change. **Evaluation and Program Planning**, 2(1), 67–90.

Fleming, T. R., & DeMets, D. L. (1996). Surrogate end points in clinical trials: Are we being misled? **Annals of Internal Medicine**, 125(7), 605–613.

Goodhart, C. A. E. (1975). Problems of monetary management: The U.K. experience. In **Papers in Monetary Economics**, Vol. 1. Sydney: Reserve Bank of Australia.

Janeway, C. A. (1989). Approaching the asymptote? Evolution and revolution in immunology. **Cold Spring Harbor Symposia on Quantitative Biology**, 54, 1–13.

MacKenzie, D. (2006). **An Engine, Not a Camera: How Financial Models Shape Markets**. Cambridge: MIT Press.

Nelson, R. R., & Winter, S. G. (1982). **An Evolutionary Theory of Economic Change**. Cambridge: Harvard University Press.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research. **Journal of Applied Psychology**, 88(5), 879–903.

Polanyi, M. (1958). **Personal Knowledge**. London: Routledge.

Polanyi, M. (1966). **The Tacit Dimension**. Garden City: Doubleday.

Strathern, M. (1997). “Improving ratings” : Audit in the British University system. **European Review**, 5(3), 305–321.

Thorndike, R. L. (1949). **Personnel Selection: Test and Measurement Techniques**. New York: Wiley.

Zinkernagel, R. M., & Doherty, P. C. (1974). Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. **Nature**, 248, 701–702.

Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. **Organization Science**, 13(3), 339–351.

沈自尹等：肾阳虚证的临床与实验研究系列报告，上海，一九五九年起。

陈可冀等：血瘀证与活血化瘀研究系列报告。

《中药新药临床研究指导原则》，一九九三年版及其后续修订（含证候疗效评价部分）。

《中医临床诊疗术语·证候部分》（GB/T 16751.2），一九九七年版及其后续修订。

朱文锋：《证素辨证学》，二〇〇八年。

附录

附录一 九个读数速查

读数	定义	分母类型	型	成本档	方向
让层比	本次判定认定成立的层数中，本次处置未予针对的层数所占比例	项数	C	乙	双向
吞差率	$(k - m) / k$ ，k 为首末直接比较判出的改变数，m 为记录中登记过的改变数	次数	M	丙	越低越好
配格率	为使这一次配得上而做的调整量 ÷ 该次处置的总量	项数	D	乙	双向
留底率	复核所依据的证据项中，取自首次处方之前已固定或结构上不受其影响者所占比例	项数	D	甲	越高越好
存异率	一次复核中出现过的分歧条目里，在归档文件中仍可检索到技术内容者所占比例	事件数	M	丙	越高越好
弃判率	全部判定输出中，记为“不适用／无从判定”且带下游处置路径的次数所占比例	次数	D	甲	双向
随程率	最近 N 次对象归属变更中，由程序或标准变更引起	次数	C	乙	双向

	者所占比例				
消参率	(原始参数数 - 可唯一确定的参数组合数) ÷ 原始参数数	结构量	M	丙	越低越好
同据率	判定依据项集合与效果判定依据项集合的交集, 占判定依据项集合的比例	项数	D	甲	越低越好

型：D 直接审计 | C 重构编码 | M 模型依赖。**成本档**：甲=一个下午翻现成文件；乙=一周至一月须重新编码；丙=须设计、埋点或建模。**四个双向量没有“越靠近某值越好”的说法，只有一个区间，而本书给不出这个区间（第十三章第三节）。**

附录二 二十七问

九句判据各拆为三问：**存在**（有没有这个字段）／**内容**（这一次填了什么）／**频次**（过去一段时期填过几次）。三问不可省，各自证否不同的东西。三条编码规则见第十二章第四节：字段须在被判定者那一份记录上；可选填并入“无”；只记有无，不记内容质量。

一·让层（第一章）

1-1 判定记录中，是否存在一个字段，记录本次认定成立而本次不予处置的病机层？

1-2 最近一次判定中，该字段填写的内容是什么？（原文抄录／为空／字段不存在）

1-3 过去十二个月的判定记录中，该字段非空的份数占比是多少？

二·吞差（第二章）

2-1 记录制度中，是否存在一次与首诊记录直接并排的比较（而非仅与上一次比较）？

2-2 最近一次这样的比较发生在何时，结论是什么？

2-3 过去十二个月中，证型改变被正式登记过几次？

三·配格（第三章）

3-1 处方记录中，是否存在一个字段，记录基础方之外各味药所针对的个体特征？

3-2 最近一张方中，该字段填写的内容是什么？

3-3 过去十二个月中，该字段非空的处方占比是多少？

四·并身（第四章）

4-1 复核记录中，各项证据是否标注了其固定时点？

4-2 最近一次复核所依据的证据项中，有哪几条写于上一次开方之前？

4-3 过去十二个月的复核中，可据以计算留底率的份数占比是多少？

五·摊差（第五章）

5-1 复核归档模板中，是否存在一个专供记录未被采纳意见的字段？（是／否／有自由文本区但不与结论同屏）

5-2 最近一次复核归档件中，该字段填写的内容是什么？

5-3 过去十二个月的复核归档件中，该字段非空的份数占比是多少？

六·填判（第六章）

6-1 判定结论表上，是否存在“不适用／无从判定”这一可填的格？该选项是否附带下游处置路径？

6-2 最近一次被填入该格的判定，其下游处置是什么？

6-3 过去十二个月中，该格被填过几次？占全部判定输出的比例是多少？

七·随程（第七章）

7-1 分类标准的修订说明中，是否载明本次修订导致的既往病例归类改变比例？

7-2 最近一次分类变更的当日记录中，是否出现了新的观察项？

7-3 最近 N 次分类变更中，有几次的当日记录里没有出现新的观察项？

八·消参（第八章）

8-1 判定所依据的观测项清单是否可得？是否做过结构可辨识性分析？

8-2 这一次判定所依据的观测项中，哪些对始终同时出现、同时缺失？

8-3 过去 N 次记录中，此类共现对的数目是多少？

九·同据（第九章）

9-1 研究方案或诊疗规范中，证候诊断所依据的条目表与证候疗效评价所依据的条目表，是否分别列出？

9-2 两份条目表的交集有几项？回路外项有几项？

9-3 过去十二个月内本机构参与的研究中，分别列出这两份表的方案占比是多少？

附录三 九条赌注

各章末尾各有一条写死日期的赌注，以及各自的“什么不算命中”。此表只列日期与要点，条款以各章原文为准。**九条赌注测的是制度扩散，不是理论真伪——赌注落空只取消相应的实践主张，不取消该章的诊断主张。**

章	期限	赌的是
一 让层	2035-12-31	出现一个独立于兼证、可被下一次就诊带出、且进入疗效评价分母调整的“本次未处置病机层”字段
二 吞差	2036-12-31	出现要求与首诊记录直接并排比较的记录规范
三 配格	2037-12-31	出现要求记录方外加减所针对的个体特征的处方规范
四 并身	2033-12-31	出现要求标注复核证据项固定时点的复核规范
五 摊差	2033-12-31	出现“未采纳意见”字段的归档要求
六 填判	2034-12-31	出现“无从判定”这一可填选项并附带下游处置路径
七 随程	2033-12-31	至少一部国家级或行业级证候诊断标准，在修订说明中载明本次修订导致的既往病例归类改变比例，并公布新旧标准之间可回溯的归类映射规则
八 消参	2034-12-31	至少一份报告规范、行业指南或标准修订说明，要求在报告任何参数估计或权重之前先报告结构可辨识性分析结果
九 同据	2034-12-31	至少一份报告规范、注册要求

		或指南制定手册，明文要求把证候诊断所依据的条目与证候疗效评价所依据的条目分别列出，并报告二者的重合项数
--	--	---

三条共同的排除条款：仅出现在方法学论文、社论或综述中而未进入任何规范、注册要求或制定手册者，不算命中；仅作出方向性倡议（如“鼓励采用客观指标”）而未指明具体字段者，不算命中；字段设立而无任何一份归档件实际填写者，不算命中。

附录四 全书三条证伪条件

其一。若九个读数在同一批对象上采齐之后，两两相关系数普遍高于 0.8，则九章讲的是同一样东西的九种说法，本书应当并章、变薄。检验方案与五对预注册见第十一章第五节；并章的统一规则：任一对高于 0.8 必须并章，不得以“侧重不同”为由保留两个名字。

其二。若在一批已把九栏立起来、九个读数均落在合理区间的机构里，判定质量在两年内没有可测的改善，则本书的实践主张失败。此条不撤销诊断主张，但书名应当改掉。

其三。若另一组研究者在不知道本书结论的情况下，用另外三组未经预选的学科重跑这九个问题，得到的不是“被吸收的量”而是别的形状，则本书的三乘三设计只是一次事后命名。**本书自己无法执行这一条**，因为执行它的条件就是不知道本书。

另有第十三章第五节的四模型比较：若可加模型胜出，第四编整个作废，本书应被拆回九篇独立论文，“结算学”这个名字应当被撤销。

附录五 九章的生成工序

九章由同一道工序写成。此处把工序摊开，一是为了让另一组研究者能够按它重跑（附录四其三），二是因为本书主张的正是“发生过的动作应当留痕”，而一本书自己的生成动作没有理由例外。

第一步：取景。 找一个每天都在发生、而没有任何一方觉得需要解释的场景。判据是：若把这个场景讲给行内人听，他的第一反应是“这有什么好说的”。

第二步：请三家。 请三个互不通约的学科传统各出一条最强硬的答法。**要求是三条互相取消前提，而不是三个角度。** 检验：三条中至少有一对，必须构成“一门学问的日课正是另一门诊断为病根的那件事”。

第三步：挖地基。 找出三条答法共同站着的那块地——一件三家都没说出口、因为在它们看来根本不需要说的假定。

第四步：自毁材料。 不从外部找反例，而是从三家中的某一家自己已经掌握、却从未往这个方向使用的成果里，找出推翻三家共同假定的那一件。九章各自的自毁材料：保护基（化学）／不可分辨关系的不传递（心理物理学）／分组选配与配对件（工程学）／无标准物质时的当轮稳健共识（分析化学）／基本常数平差的扩展因子（物理学）／聚类算法必然分配标签（细胞学）／被认证的程序反过来规定下一批标准物质（分析化学）／参数估计不收敛而闭环跟踪误差趋零（工程学）／胸腺阳性选择与组织相容性限制（免疫学）。

第五步：命名。 为那件“在三套账上都不设字段”的东西起一个名字，并写明它不是什么——与最近的几个现成说法逐一划界。

第六步：给读数。 一个跨领域量纲不同而比率相同的数，且必须可清点。

第七步：给判据。 一句不含情态词的问话，答案必须去查记录，不能靠回想。

第八步：削弱。 三处反过来削弱本章的地方、适用范围的收窄、可以让本章翻车的观测、本章解释不了的洞、本章自己的读数、反噬预言，以及一条写死日期的赌注与“什么不算命中”。

这道工序有一个本书无法自检的风险，已写在第十章第二节：**它本身可能就是一台“生产被吸收的量”的机器**——把它对准任何题目，都会产出一样形状相同的东西。附录四其三是唯一能检验这一点的设计。

后记

三笔欠账

第一笔。九句判据里最便宜的三条——把诊断标准的条目表与疗效评价的条目表并排放数交集、数一段时期的判定输出里有几次记为无从判定、看一份复核所依据的证据项各自的日期——一个下午可以做完。本书一次也没有做。这一笔在导论第九节、第十一章第八节、结语第三节各写过一次，此处是第四次，因为它是本书与“一本可以被相信的书”之间距离最短的一处。

第二笔。二十七问的编码者一致性没有做。两组人各自独立编码一百份既有归档记录即可，不需要任何机构改变流程。在做出来之前，二十七问只能用于探索，本书自己给出的伦理第三条因此适用于本书自己。

第三笔。本书说得出该改什么，说不出该由谁改。病历书写规范与临床研究报告规范的修订权，不在任何一位读过本书的临床医生手里，也不在作者手里。本书的处方落在自己第六章所指认的那种位置上——一个本该有归属的地方，没有人被指派。

承重文献的核验状态

本书的文献分三类：**承重类**（本书某一步论证直接依赖其结论）、**划界类**（用于与近邻区分）、**背景类**（提供语境）。第一类须逐条回原文核验并留存核验表（引文、来源、是否一致、页码、DOI、核验人）。本书尚未完成这一轮核验，书末所列为文献线索与择要著录，不是已核验清单。引用者请核对原文，若有出入以原文为准。

第一至三章的文献以线索体列出（按出现次序的可核查方向），第四至九章以择要著录体列出。两种体例的差异是有意保留的——它反映了三组学科各自的文献习惯。若日后出国际版，应统一为一种著录规范。

人机分工声明

本书的问题、三组学科的指定、九个题目的划分、以及全部成文要求，由作者提出。材料梳理、论证组织与文字撰写，由作者与人工智能在多轮往复中协作

完成。全部判断、每一处削弱的接受、适用范围的收窄、赌注条款与“什么不算命中”的逐条写死，由作者定稿并负责。

九章原各自附有一段同样内容的声明，装书时合并为此一处，不逐章重复。

各章第十六节（或相应位置）所列的最强竞争解释，是在写作过程中被专门找出来自我攻击的，不是事后补上的。各章“版本”一行所记的削减与改写，均为成稿过程中实际发生的修改，其中数处使本书的主张变弱——例如第一章的适用范围由“任何判定”收窄为“处置手段选择性低于对象分节精度的系统”，第二章的比率定义由一个不可清点的形式改为可清点的形式，第三章的一条限制（配格率不得设为考核指标）是在接受了一处削弱之后才写下的，此前本书原打算主张配格率越高越好。

这些修改的技术内容没有被保留。按本书第五章自己的读数，**本书的存异率为零**。这是本书自己指认的那件事发生在本书自己身上，且这一次没有借口——加一个附录就能解决，本书没有加。

一句

九章合起来只主张一件事：每一次判定都做了一些事，账上留下的是一个证名。

其余的话已经在结语里说完了。

胡敏